Journal of Statistical Planning and Inference 94 (2001) 25–36

journal of statistical planning and inference

www.elsevier.com/locate/jspi

# Some tests for the equality of covariance matrices

James R. Schott *

*Department of Statistics, University of Central Florida, Orlando, FL 32816-2370, USA*

Received 13 January 2000; received in revised form 15 June 2000; accepted 10 August 2000

**Abstract**

A Wald statistic, which is asymptotically equivalent to the likelihood ratio criterion, is obtained for the test of the equality of covariance matrices. A more general Wald statistic is constructed under the assumption of elliptical distributions, and the comparison of these two statistics sheds some light on the asymptotic performance of the likelihood ratio test. In particular, we find that the likelihood ratio test is liberal for nonnormal elliptical populations with positive kurtosis and conservative for nonnormal elliptical populations with negative kurtosis. Further, the likelihood ratio test cannot be adjusted by a scalar multiple so as to retain its asymptotic chi-squared distribution over the class of elliptical distributions. A Wald test, appropriate for more general populations, is also obtained. 

*MSC*: 62H15

*Keywords*: Elliptical distribution; Kurtosis; Wald statistic

## 1. Introduction

Many multivariate analyses of grouped data require a test of the equality of covariance matrices. In some applications, the comparison of covariance matrices is of primary interest. An example of this is the common principal components analysis (Krzanowski, 1979, Flury, 1988, Schott, 1991). In this case, the test for equal covariance matrices is used as an initial test to determine whether any further analysis is necessary. A second general application of this test of equality is to check the assumption of equal covariance matrices that applies to many of the standard multivariate analyses such as multivariate analysis of variance and linear discriminant analysis.

Suppose that we have $k$ groups with the $i$th group having an $m$-variate multivariate distribution with finite fourth moments, mean vector $\mu_i$ and covariance matrix $\Omega_i$ which is positive definite. In addition, we assume that we have independent random samples

* Tel.: +1-407-823-2269.

0378-3758/01/$ - see front matter 
PII: S0378-3758(00)00209-3

from the $k$ groups, with a sample of size $N_i = n_i + 1$ from the $i$th group, from which we compute the usual unbiased sample covariance matrices, $S_1, \dots, S_k$; that is,

$$S_i = \frac{1}{n_i} \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)',$$

where $x_{i1}, \dots, x_{iN_i}$ is the sample from the $i$th group and $\bar{x}_i = N_i^{-1} \sum_{j=1}^{N_i} x_{ij}$. A test of the null hypothesis

$$\mathrm{H}_0\colon \Omega_1 = \cdots = \Omega_k = \Omega,$$

for unknown common covariance matrix $\Omega$, versus the general alternative is most commonly based on the statistic

$$M = n \log|S| - \sum_{i=1}^{k} n_i \log|S_i|,$$

where

$$S = \sum_{i=1}^{k} \gamma_i S_i,$$

$\gamma_i = n_i/n$, and $n = n_1 + \cdots + n_k$. This test, which is known as Bartlett's test, is based on the likelihood ratio criterion after modification so as to yield an unbiased test (see, for example, Muirhead, 1982, Section 8.2). If each group has a multivariate normal distribution and $\mathrm{H}_0$ holds, then the asymptotic distribution of $M$ is chi-squared with $v = (k-1)m(m+1)/2$ degrees of freedom.

It is well known that this likelihood ratio test is very sensitive to violations of the normality assumption, and so other more robust procedures have been proposed. Tiku and Balakrishnan (1985) developed a test for the equality of two covariance matrices by adapting a robust test for two mean vectors. O'Brien (1992) obtained robust tests by generalizing Levine's test for the equality of variances. Zhang and Boos (1992) proposed a bootstrap procedure to estimate critical values to be used in conjunction with Bartlett's statistic. In this paper, we develop several Wald tests for $\mathrm{H}_0$.

## 2. Some Wald tests

In this section, we consider the construction of Wald statistics for testing $\mathrm{H}_0$ under various assumptions regarding the underlying distributions possessed by the $k$ populations. In the simplest case, we assume that each population has a multivariate normal distribution. The resulting statistic can be used to investigate the asymptotic behavior of the likelihood ratio criterion, $M$, since the two statistics are asymptotically equivalent. When it is assumed that each population has an elliptical distribution, we obtain a statistic which is computationally nearly as simple as the normal-theory statistic while having broader applicability. The comparison of these two Wald statistics gives some insight into the asymptotic performance of the likelihood ratio test for nonnormal populations. Finally, we consider the most general case in which we assume that the distributions, which may differ from population to population, have finite fourth moments.

### 2.1. Multivariate normal populations

A Wald statistic can be easily constructed by comparing the sample covariance matrices. For instance, we can utilize the vector of differences $v=(v_1',\dots,v_{k-1}')'$, where $v_i=\mathrm{vec}(S_i-S_k)$. Here, $\mathrm{vec}(S_i-S_k)$ represents the $m^2$-dimensional vector formed by stacking the columns of $S_i-S_k$. Since our $k$ samples are independent, the sample covariance matrices are independently distributed and, in particular, $n_i^{1/2}\mathrm{vec}(S_i-\Omega_i)$ has an asymptotic normal distribution with mean vector 0 and covariance matrix which we will denote by $\Psi_i$. When each of the $k$ groups has a multivariate normal distribution, $\Psi_i=(I_{m^2}+K_{mm})(\Omega_i\otimes\Omega_i)$ and so if $\mathrm{H}_0$ holds, the asymptotic covariance matrix $\Phi_1$ of $n^{1/2}v$ is given by

$$\Phi_1=(D^{-1}+\gamma_k^{-1}1_{k-1}1_{k-1}')\otimes\{(I_{m^2}+K_{mm})(\Omega\otimes\Omega)\}, \tag{1}$$

where $D=\mathrm{diag}(\gamma_1,\dots,\gamma_{k-1})$, $1_{k-1}$ is the $(k-1)\times 1$ vector of 1's, and $K_{mm}$ is a commutation matrix which satisfies the commutative relationship $K_{mm}(A\otimes B)=(B\otimes A)K_{mm}$ for any $m\times m$ matrices $A$ and $B$ (see, for example, Magnus, 1988, Chapter 3). The form of the Wald statistic computed from $v$, which we will denote by $T_1$, is given in the following theorem.

**Theorem 1.** *If* $\mathrm{H}_0$ *holds and our k populations have multivariate normal distributions, then the statistic*

$$T_1=\frac{n}{2}\left\{\sum_{i=1}^{k}\gamma_i\,\mathrm{tr}(S_iS^{-1}S_iS^{-1})-\sum_{i=1}^{k}\sum_{j=1}^{k}\gamma_i\gamma_j\,\mathrm{tr}(S_iS^{-1}S_jS^{-1})\right\}$$

*has an asymptotic chi-squared distribution with degrees of freedom* $v=(k-1)m(m+1)/2$.

**Proof.** The Moore–Penrose generalized inverse of $\Phi_1$ is

$$\Phi_1^{+}=(D-\gamma\gamma')\otimes\left\{\frac{1}{4}(I_{m^2}+K_{mm})(\Omega^{-1}\otimes\Omega^{-1})\right\},$$

where $\gamma$ is the $(k-1)\times 1$ vector $(\gamma_1,\dots,\gamma_{k-1})'$. A consistent estimator $\hat{\Phi}_1^{+}$ of $\Phi_1^{+}$ can be obtained by replacing $\Omega$ in the expression for $\Phi_1^{+}$ by $S$. It follows from the general theory of Wald statistics (Moore, 1977) that $T_1=nv'\hat{\Phi}_1^{+}v$ has an asymptotic chi-squared null distribution when $n_i$ converges to infinity in such a way that $\gamma_i\to\gamma_i^*\in(0,1)$ for $i=1,\dots,k$. The degrees of freedom, $v$, for this chi-squared distribution is given by

$$v=\mathrm{rank}(\Phi_1)=\mathrm{rank}(D^{-1}-\gamma_k^{-1}1_{k-1}1_{k-1}')\times\mathrm{rank}\,\{(I_{m^2}+K_{mm})(\Omega\otimes\Omega)\}$$

$$=(k-1)\times m(m+1)/2.$$

Finally, using the fact that

$$\begin{aligned} v_i'(I_{m^2}+K_{mm})(S^{-1}\otimes S^{-1})v_j &= \mathrm{vec}(S_i-S_k)'(I_{m^2}+K_{mm})(S^{-1}\otimes S^{-1})\mathrm{vec}(S_j-S_k) \\ &= 2\mathrm{tr}\{(S_i-S_k)S^{-1}(S_j-S_k)S^{-1}\}, \end{aligned}$$

we find that if we let $\delta_{ij}$ denote the $(i,j)$th element of $D-\gamma\gamma'$, then

$$\begin{aligned} T_1 &= \frac{n}{4}\sum_{i=1}^{k-1}\sum_{j=1}^{k-1}\delta_{ij}v_i'(I_{m^2}+K_{mm})(S^{-1}\otimes S^{-1})v_j \\ &= \frac{n}{2}\left\{\sum_{i=1}^{k-1}\gamma_i\mathrm{tr}\{(S_i-S_k)S^{-1}(S_i-S_k)S^{-1}\}\right. \\ &\quad \left. -\sum_{i=1}^{k-1}\sum_{j=1}^{k-1}\gamma_i\gamma_j\mathrm{tr}\{(S_i-S_k)S^{-1}(S_j-S_k)S^{-1}\}\right\} \\ &= \frac{n}{2}\left\{\sum_{i=1}^{k}\gamma_i\mathrm{tr}(S_iS^{-1}S_iS^{-1})-\sum_{i=1}^{k}\sum_{j=1}^{k}\gamma_i\gamma_j\mathrm{tr}(S_iS^{-1}S_jS^{-1})\right\}. \end{aligned}$$

Theorem 1 provides an alternative to the likelihood ratio test for testing $H_0$ when we have independent random samples from normal populations. However, it is easily shown that the two tests are asymptotically equivalent; that is, $M = T_1 + o_p(1)$. Consequently, the test based on $T_1$, like that based on $M$, will be overly sensitive to violations of the normality assumption. □

### 2.2. Elliptical populations with common kurtosis parameter

By constructing a Wald statistic under the assumption of elliptical distributions, we will obtain a test which performs better than $M$ and $T_1$ outside the class of normal distributions. It is well known (Muirhead and Waternaux, 1980, Tyler, 1983) that many normal-theory tests can be adjusted by a simple scalar multiple so as to retain their asymptotic distribution in the class of elliptical distributions. Although this was claimed by Zhang and Boos (1992) to be the case for $M$, we will see that such an adjustment is not possible.

An $m\times 1$ random vector $x$ has an elliptical distribution with mean vector $\mu$ and covariance matrix $\Omega$ if its characteristic function has the form, $\phi(t) = \mathrm{e}^{it'\mu}\psi(t'\Omega t)$, where $\psi$ is a function scaled so that $\psi'(0) = 1/2$. We will assume that each of our $k$ groups has an elliptical distribution with common kurtosis parameter; that is, if $x_i$ represents a random vector having the distribution of the $i$th population, we assume that $\kappa_1 = \cdots = \kappa_k = \kappa$, where

$$\kappa_i = \frac{E[\{\mathrm{e}_l'(x_i-\mu_i)\}^4]}{3(\sigma_{ll}^i)^2} - 1, \tag{2}$$

$e_l$ denotes the $l$th column of $I_k$ and $\sigma^i_{ll}$ is the $(l,l)$th element of $\Omega_i$. In this case, if $\mathrm{H}_0$ holds, $\Psi_i = (1+\kappa)(I_{m^2}+K_{mm})(\Omega\otimes\Omega)+\kappa\mathrm{vec}(\Omega)\mathrm{vec}(\Omega)'$ and so the asymptotic covariance matrix of $n^{1/2}v$ can be expressed as

$$\Phi_2 = (D^{-1}+\gamma_k^{-1}1_{k-1}1'_{k-1})\otimes\{(1+\kappa)(I_{m^2}+K_{mm})(\Omega\otimes\Omega)+\kappa\,\mathrm{vec}(\Omega)\mathrm{vec}(\Omega)'\}. \tag{3}$$

The fact that the covariance matrix given in (3) is not simply a scalar multiple of the covariance matrix given in (1) implies that $T_1$ cannot be adjusted by a scalar mutiple so as to retain its asymptotic null distribution in the class of elliptical distributions.

We will not be able to construct a Wald test that is valid for all elliptical distributions since the rank of $\Phi_2$ depends on $\kappa$. Now for all elliptical distributions, $\kappa \geqslant -2/(m+2)$. This inequality is easily obtained by using the stochastic representation (Fang et al., 1990), $x=\mu+rTu$, for a random vector $x$ having an elliptical distribution with mean vector $\mu$ and covariance matrix $\Omega$. Here $r$ is a nonnegative random variable independent of $u$ which has a uniform distribution on the unit sphere in $R^m$, and $TT'=\Omega$. The kurtosis parameter for this elliptical distribution simplifies to

$$\kappa = \frac{mE(r^4)}{(m+2)\{E(r^2)\}^2}-1 \tag{4}$$

and the desired inequality follows from the fact that $E(r^4)\geqslant\{E(r^2)\}^2$ for any random variable $r$. If $\kappa>-2/(m+2)$, $\mathrm{rank}(\Phi_2)=(k-1)m(m+1)/2$ while $\kappa=-2/(m+2)$, which is the case for the uniform distribution on the $m$-dimensional unit sphere, implies that $\mathrm{rank}(\Phi_2)=(k-1)\{m(m+1)/2-1\}$. The form of our Wald statistic, $T_2$, for elliptical populations with common kurtosis parameter $\kappa > -2/(m+2)$ is given in the next theorem.

**Theorem 2.** *Suppose that each population has an elliptical distribution with common kurtosis parameter* $\kappa > -2/(m+2)$. *Let* $\hat\delta_1=(1+\hat\kappa)^{-1}$ *and* $\hat\delta_2=\hat\kappa/[2(1+\hat\kappa)\{2(1+\hat\kappa)+m\hat\kappa\}]$, *where* $\hat\kappa$ *is a consistent estimator of* $\kappa$. *Then, under* $\mathrm{H}_0$, *it follows that*

$$T_2 = n\left(\sum_{i=1}^k\left\{\frac{1}{2}\hat\delta_1\gamma_i\mathrm{tr}(S_iS^{-1}S_iS^{-1})-\hat\delta_2\gamma_i\mathrm{tr}(S_iS^{-1})^2\right\}\right.$$
$$\left.-\sum_{i=1}^k\sum_{j=1}^k\left\{\frac{1}{2}\hat\delta_1\gamma_i\gamma_j\mathrm{tr}(S_iS^{-1}S_jS^{-1})-\hat\delta_2\gamma_i\gamma_j\mathrm{tr}(S_iS^{-1})\mathrm{tr}(S_jS^{-1})\right\}\right)$$

*has an asymptotic chi-squared distribution with* $v=(k-1)m(m+1)/2$ *degrees of freedom.*

**Proof.** It is easily verified that when $\kappa > -2/(m+2)$

$$\Phi_2^+ = (D-\gamma\gamma')\otimes\left\{\frac{1}{4}\delta_1(I_{m^2}+K_{mm})(\Omega^{-1}\otimes\Omega^{-1})-\delta_2\,\mathrm{vec}(\Omega^{-1})\mathrm{vec}(\Omega^{-1})'\right\},$$

where $\delta_1=(1+\kappa)^{-1}$ and

$$\delta_2 = \frac{\kappa}{2(1+\kappa)\{2(1+\kappa)+m\kappa\}}.$$

A consistent estimator $\hat{\Phi}_2^+$ can then be obtained by replacing $\Omega$ by $S$ and $\kappa$ by any consistent estimator $\hat{\kappa}$. As in the previous proof, it then follows that $T_2 = nv'\hat{\Phi}_2^+ v$ converges in distribution to a chi-squared with $v = \text{rank}(\Phi_2) = (k-1)m(m+1)/2$ degrees of freedom. Letting $\hat{\delta}_1$ and $\hat{\delta}_2$ denote the estimators obtained by substituting $\hat{\kappa}$ for $\kappa$ in the expressions for $\delta_1$ and $\delta_2$, $T_2$ can be expressed as

$$\begin{aligned}
T_2 = n\Bigg( &\sum_{i=1}^{k-1}\left[\frac{1}{2}\hat{\delta}_1\gamma_i \text{tr}\{(S_i - S_k)S^{-1}(S_i - S_k)S^{-1}\} - \hat{\delta}_2\gamma_i \text{tr}\{(S_i - S_k)S^{-1}\}^2\right] \\
&- \sum_{i=1}^{k-1}\sum_{j=1}^{k-1}\left[\frac{1}{2}\hat{\delta}_1\gamma_i\gamma_j \text{tr}\{(S_i - S_k)S^{-1}(S_j - S_k)S^{-1}\}\right. \\
&\left.-\hat{\delta}_2\gamma_i\gamma_j \text{tr}\{(S_i - S_k)S^{-1}\}\text{tr}\{(S_j - S_k)S^{-1}\}\right]\Bigg) \\
= n\Bigg( &\sum_{i=1}^{k}\left\{\frac{1}{2}\hat{\delta}_1\gamma_i \text{tr}(S_iS^{-1}S_iS^{-1}) - \hat{\delta}_2\gamma_i \text{tr}(S_iS^{-1})^2\right\} \\
&- \sum_{i=1}^{k}\sum_{j=1}^{k}\left\{\frac{1}{2}\hat{\delta}_1\gamma_i\gamma_j \text{tr}(S_iS^{-1}S_jS^{-1}) - \hat{\delta}_2\gamma_i\gamma_j \text{tr}(S_iS^{-1})\text{tr}(S_jS^{-1})\right\}\Bigg),
\end{aligned}$$

and so the proof is complete. □

By comparing $T_1$ and $T_2$, we can get some insight into the asymptotic performance of $T_1$ when sampling from nonnormal elliptical distributions. Note that

$$\Phi_2^+ = \delta_1\Phi_1^+ - \delta_2\{(D - \gamma\gamma') \otimes \text{vec}(\Omega^{-1})\text{vec}(\Omega^{-1})'\},$$

so that $T_2 = \delta_1 T_1 - \delta_2 T_* + o_p(1)$, where the random variable

$$T_* = nv'\{(D - \gamma\gamma') \otimes \text{vec}(\Omega^{-1})\text{vec}(\Omega^{-1})'\}v$$

is nonnegative with probability one. Consequently, we may write $T_1$ as

$$T_1 = \delta_1^{-1}(T_2 + \delta_2 T_*) + o_p(1). \tag{5}$$

When $\kappa > 0$, we have $\delta_1 < 1$ and $\delta_2 > 0$ and so asymptotically $T_1$ is stochastically larger than $T_2$, that is, stochastically larger than a chi-squared random variable with $v$ degrees of freedom. In other words, when $\kappa > 0$, the test based on $T_1$ or $M$ will yield significance levels above the nominal significance level. This confirms the liberal nature of the likelihood ratio test that others (Hopkins and Clay, 1963; Layard, 1974; Olson, 1974) have demonstrated empirically. However, this test can also be conservative. If $-2/(m+2) < \kappa < 0$, then $\delta_1 > 1$ and $\delta_2 < 0$, and from this and (5) it follows that asymptotically $T_1$ is stochastically smaller than a chi-squared random variable with $v$ degrees of freedom.

### 2.3. More general settings

In this section, we obtain Wald statistics under less restrictive assumptions regarding the distributions for our $k$ groups. In particular, we consider the most general case in which each population has some distribution with finite fourth moments. As a special case, we consider populations having elliptical distributions with different kurtosis parameters. The resulting Wald statistics for these two cases will be denoted by $T_4$ and $T_3$, respectively. We will need to assume that the covariance matrix $\Psi_i$ has a rank of $m(m+1)/2$ for each $i$. Since $\Psi_i = M_{4i} - \text{vec}(\Omega_i)\text{vec}(\Omega_i)'$, where

$$M_{4i} = E[(x_i - \mu_i)(x_i - \mu_i)' \otimes (x_i - \mu_i)(x_i - \mu_i)'],$$

it follows that this last assumption is equivalent to the condition that

$$\text{var}\{(x_i - \mu_i)'B(x_i - \mu_i)\} > 0 \tag{6}$$

for any $m \times m$ matrix $B$ for which $B \neq -B'$.

We will find it more convenient to work with the vectors $v(S_i)$ instead of the vectors $\text{vec}(S_i)$, where $v(S_i)$ denotes the $m(m+1)/2 \times 1$ vector which is obtained from $\text{vec}(S_i)$ by eliminating all elements above the diagonal of $S_i$. Let $H$ be any $m(m+1)/2 \times m^2$ matrix for which $H\,\text{vec}(A) = v(A)$ whenever $A$ is an $m \times m$ symmetric matrix; for instance, one choice for $H$ is the elimination matrix $L_m$ which has $e_i' \otimes e_j'$ as its $\{(i-1)m + j - (1/2)i(i-1)\}$th row, $j \geqslant i$ (see, for example, Magnus, 1988, Chapter 5). The asymptotic distribution of $n_i^{1/2}v(S_i)$ is normal with a mean vector $v(\Omega_i)$ and covariance matrix $H\Psi_i H'$ which is nonsingular due to condition (6). Thus, a test of $\text{H}_0$ essentially reduces to a one-way multivariate analysis of variance with unequal covariance matrices. Consequently, our next result follows immediately from James (1954). This Wald statistic requires a consistent estimator $\hat{\Psi}_i$ which can be obtained from the expression for $\Psi_i$ by replacing $\Omega_i$ by $S$ and $M_{4i}$ by any consistent estimator $\hat{M}_{4i}$.

**Theorem 3.** *Define* $W_i$ *and* $W$ *as*

$$W_i = \gamma_i(H\hat{\Psi}_i H')^{-1}, \quad W = \sum_{i=1}^{k} W_i.$$

*Then under* $\text{H}_0$ *and condition* (6), *the statistic*

$$T_4 = n\left(\sum_{i=1}^{k} v(S_i)'W_i v(S_i) - \sum_{i=1}^{k}\sum_{j=1}^{k} v(S_i)'W_i W^{-1} W_j v(S_j)\right) \tag{7}$$

*has an asymptotic chi-squared distribution with degrees of freedom* $v = (k-1)m(m+1)/2$.

If the $i$th population has an elliptical distribution with kurtosis parameter $\kappa_i$ given in (2) and $\text{H}_0$ holds, then the covariance matrix $\Psi_i$ simplifies to

$$\Psi_i = (1+\kappa_i)(I_{m^2} + K_{mm})(\Omega \otimes \Omega) + \kappa_i\,\text{vec}(\Omega)\text{vec}(\Omega)'.$$

When $S$ and a consistent estimator $\hat{\kappa}_i$ of $\kappa_i$ are substituted into this expression to obtain an estimator $\hat{\Psi}_i$, and this is used in Theorem 3, the resulting Wald statistic has a simpler form. This is given in our final theorem.

**Theorem 4.** *Let* $\hat{\kappa}_i$ *be a consistent estimator of* $\kappa_i$ *and define* $\hat{\alpha}_i = 1/2\gamma_i(1+\hat{\kappa}_i)^{-1}$, $\hat{\alpha} = \sum \hat{\alpha}_i$, $\hat{\beta} = \sum \hat{\beta}_i$,

$$\hat{\beta}_i = -\frac{\gamma_i \hat{\kappa}_i}{2(1+\hat{\kappa}_i)\{2(1+\hat{\kappa}_i)+m\hat{\kappa}_i\}},$$

$$\hat{\rho} = -\frac{\hat{\beta}}{\hat{\alpha}(\hat{\alpha}+m\hat{\beta})},$$

$$\hat{\tau}_{ij} = \hat{\alpha}^{-1}\hat{\alpha}_i\hat{\beta}_j + (\hat{\alpha}_i\hat{\rho} + \hat{\alpha}^{-1}\hat{\beta}_i + m\hat{\beta}_i\hat{\rho})(\hat{\alpha}_j + m\hat{\beta}_j).$$

*Then, under* $\mathrm{H}_0$ *and the assumption that the ith population has an elliptical distribution with kurtosis parameter* $\kappa_i > -2/(m+2)$, *it follows that*

$$T_3 = n\left(\sum_{i=1}^{k}\{\hat{\alpha}_i \operatorname{tr}(S_iS^{-1}S_iS^{-1}) + \hat{\beta}_i \operatorname{tr}(S_iS^{-1})^2\}\right.$$
$$\left. - \sum_{i=1}^{k}\sum_{j=1}^{k}\{\hat{\alpha}^{-1}\hat{\alpha}_i\hat{\alpha}_j \operatorname{tr}(S_iS^{-1}S_jS^{-1}) + \hat{\tau}_{ij}\operatorname{tr}(S_iS^{-1})\operatorname{tr}(S_jS^{-1})\}\right)$$

*has an asymptotic chi-squared distribution with* $v = (k-1)m(m+1)/2$ *degrees of freedom.*

**Proof.** In proving the result, we utilize the duplication matrix $D_m$ (Magnus, 1988, Chapter 4), its Moore–Penrose inverse $D_m^+$, and associated properties such as $D_m^+\operatorname{vec}(S)= v(S)$, $D_m v(S) = \operatorname{vec}(S)$, $D_m^+D_m = I_{m(m+1)/2}$, and $2D_mD_m^+ = I_{m^2} + K_{mm}$. The matrix $D_m$ gets its name from the fact that it duplicates the necessary elements of $v(S)$ so as to produce $\operatorname{vec}(S)$. It is readily shown that

$$W_i = \gamma_i(D_m^+\hat{\Psi}_iD_m^{+'})^{-1}$$
$$= D_m'\left\{\frac{1}{2}\hat{\alpha}_i(I_{m^2}+K_{mm})(S^{-1}\otimes S^{-1}) + \hat{\beta}_i\operatorname{vec}(S^{-1})\operatorname{vec}(S^{-1})'\right\}D_m.$$

Consequently, $W = D_m'\{\frac{1}{2}\hat{\alpha}(I_{m^2}+K_{mm})(S^{-1}\otimes S^{-1}) + \hat{\beta}\operatorname{vec}(S^{-1})\operatorname{vec}(S^{-1})'\}D_m$, while its inverse is given by

$$W^{-1} = D_m^+\left\{\frac{1}{2}\hat{\alpha}^{-1}(I_{m^2}+K_{mm})(S\otimes S) + \hat{\rho}\operatorname{vec}(S)\operatorname{vec}(S)'\right\}D_m^{+'}.$$

The result now follows by substituting these expressions for $W_i$ and $W^{-1}$ in (7) and simplifying.

## 3. Simulation results

Some simulation results were obtained so as to assess the effectiveness of the asymptotic chi-squared distribution in approximating the actual null distributions of $T_1, T_2, T_3$

and $T_4$ for finite sample sizes. The true type I error probability was estimated for $k=3$ and $m=2,5$. For simplicity, we used equal sample sizes. In each case, the nominal significance level was 0.05 and the estimated significance level was based on 1000 simulations.

The statistics $T_2$ and $T_3$ require estimates of kurtosis parameters while $T_4$ requires an estimate of $M_{4i}$, and we found that the choice of these estimates can have a significant impact on the significance levels for small samples. A simple consistent estimator of $\kappa_i$ can be obtained by replacing $\xi_l^i = E[\{e_l'(x_i - \mu_i)\}^4]$ and $(\sigma_{ll}^i)^2$ in (2) by simple consistent estimators. We used a slightly more complicated estimator of $\kappa_i$ defined by

$$\hat{\kappa}_i = \frac{1}{3m}\sum_{l=1}^{m}\frac{z_l^i}{w_l^i} - 1,$$

where

$$z_l^i = \frac{\sum_{j=1}^{N_i}\{e_l'(x_{ij} - \bar{x}_i)\}^4 - 6(s_{ll}^i)^2}{N_i - 4},$$

$$w_l^i = \frac{N_i}{(N_i - 1)}\left\{(s_{ll}^i)^2 - \frac{z_l^i}{N_i}\right\},$$

and $x_{i1},\dots,x_{iN_i}$ denotes the sample from the $i$th group while $\bar{x}_i$ is the corresponding sample mean vector. It is easily shown that $E(z_l^i) = \xi_l^i + \mathrm{O}(N_i^{-2})$ and $E(w_l^i) = (\sigma_{ll}^i)^2 + \mathrm{O}(N_i^{-2})$. The estimators $\hat{\kappa}_1,\dots,\hat{\kappa}_k$ were used in $T_3$ while $\hat{\kappa} = (\hat{\kappa}_1 + \cdots + \hat{\kappa}_k)/k$ was used in $T_2$. Similarly, we are led to an estimator of $M_{4i}$ given by

$$\hat{M}_{4i} = \frac{1}{(N_i - 4)}\left(\sum_{j=1}^{N_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \otimes (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \right.$$
$$\left. -2\{\mathrm{vec}(S)\,\mathrm{vec}(S)' + (I_{m^2} + K_{mm})(S \otimes S)\}\right).$$

The simulation results are given in Table 1. The different portions of the table, (a)–(f), correspond to six different settings used for the population distributions. It is not surprising that our most general test, the one based on $T_4$, exhibits very slow convergence to the asymptotic distribution and, in fact, for most of the cases considered has very inflated significance levels. The poor performance of this procedure, which is much more pronounced for the larger value of $m$, is due to the fact that it requires estimates of the general fourth-order moment matrices $\Psi_i$.

Table 1(a) contains results for normal populations. As would be expected, the normal-theory test, $T_1$, performs the best although $T_2$ and $T_3$ performly fairly well except for the smallest sample size. In Table 1(b) each population has a multivariate $t$ distribution with 5 degrees of freedom, while for Table 1(c) two of the populations are normal and the third one is multivariate $t$. In both cases, $T_1$ consistently yields inflated significance levels while $T_2$ and $T_3$ produce reasonable results for sample sizes of 20 or more. In Table 1(c), $T_2$ performs slightly better than $T_3$ even though this is a situation in which one would expect $T_3$ to be superior. The kurtosis parameters

Table 1
Estimated significance levels when $k=3$ and the nominal significance level is 0.05

| $m=2$ | | | | | $m=5$ | | | |
|---|---|---|---|---|---|---|---|---|
| $N_i$ | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_1$ | $T_2$ | $T_3$ | $T_4$ |
| (a) Normal populations | | | | | | | | |
| 10 | 0.049 | 0.144 | 0.195 | 0.045 | 0.049 | 0.146 | 0.165 | 0.448 |
| 20 | 0.049 | 0.068 | 0.084 | 0.080 | 0.039 | 0.081 | 0.095 | 0.768 |
| 30 | 0.049 | 0.054 | 0.065 | 0.077 | 0.052 | 0.072 | 0.080 | 0.519 |
| 50 | 0.044 | 0.057 | 0.067 | 0.076 | 0.051 | 0.063 | 0.065 | 0.336 |
| (b) Multivariate $t$ populations | | | | | | | | |
| 10 | 0.208 | 0.107 | 0.150 | 0.039 | 0.365 | 0.062 | 0.090 | 0.509 |
| 20 | 0.318 | 0.078 | 0.081 | 0.050 | 0.592 | 0.039 | 0.048 | 0.785 |
| 30 | 0.392 | 0.084 | 0.078 | 0.066 | 0.681 | 0.044 | 0.048 | 0.538 |
| 50 | 0.432 | 0.069 | 0.061 | 0.073 | 0.759 | 0.059 | 0.058 | 0.275 |
| (c) Normal and multivariate $t$ populations | | | | | | | | |
| 10 | 0.091 | 0.141 | 0.204 | 0.044 | 0.120 | 0.133 | 0.179 | 0.507 |
| 20 | 0.132 | 0.072 | 0.090 | 0.108 | 0.173 | 0.064 | 0.082 | 0.827 |
| 30 | 0.159 | 0.096 | 0.100 | 0.127 | 0.229 | 0.076 | 0.079 | 0.638 |
| 50 | 0.157 | 0.078 | 0.080 | 0.110 | 0.269 | 0.060 | 0.062 | 0.449 |
| (d) Normal and contaminated normal populations | | | | | | | | |
| 50 | 0.447 | 0.175 | 0.172 | 0.346 | 0.802 | 0.130 | 0.137 | 0.830 |
| 100 | 0.528 | 0.138 | 0.125 | 0.319 | 0.943 | 0.097 | 0.090 | 0.862 |
| 200 | 0.557 | 0.107 | 0.093 | 0.260 | 0.933 | 0.086 | 0.070 | 0.797 |
| 500 | 0.569 | 0.091 | 0.064 | 0.159 | 0.951 | 0.072 | 0.063 | 0.603 |
| (e) Elliptical populations with negative kurtosis | | | | | | | | |
| 10 | 0.001 | 0.164 | 0.157 | 0.073 | 0.006 | 0.175 | 0.176 | 0.497 |
| 20 | 0.000 | 0.133 | 0.146 | 0.057 | 0.007 | 0.093 | 0.116 | 0.756 |
| 30 | 0.000 | 0.106 | 0.136 | 0.058 | 0.006 | 0.080 | 0.100 | 0.520 |
| 50 | 0.000 | 0.071 | 0.095 | 0.030 | 0.003 | 0.053 | 0.072 | 0.277 |
| (f) Nonelliptical populations | | | | | | | | |
| 10 | 0.325 | 0.096 | 0.114 | 0.041 | 0.512 | 0.024 | 0.034 | 0.497 |
| 20 | 0.428 | 0.066 | 0.066 | 0.071 | 0.744 | 0.013 | 0.013 | 0.735 |
| 30 | 0.530 | 0.102 | 0.080 | 0.070 | 0.825 | 0.006 | 0.008 | 0.489 |
| 50 | 0.619 | 0.087 | 0.069 | 0.054 | 0.904 | 0.006 | 0.006 | 0.254 |

for the normal and multivariate $t$ distributions are 0 and 2, respectively. For Table 1(d), we replaced the single multivariate $t$ distribution by another distribution with even larger kurtosis. The distribution used was the contaminated normal, in particular, a mixture of $N_m(0, \alpha_1 I_m)$ and $N_m(0, \alpha_2 I_m)$ with probabilities 0.05 and 0.95, respectively. Here $\alpha_1 = 10$, $\alpha_2 = 0.5/0.95$, and so the kurtosis parameter (see, Muirhead 1982, p. 49) is 4.263. For this portion of the table, we needed to look at much larger values of $N_i$ since the large kurtosis parameter implies that one of the sample covariance matrices is converging to normality very slowly. Here we do have $T_3$ performing better than $T_2$, but the difference is slight. It seems that the test based on $T_2$ is fairly robust to the violation of the assumption of common kurtosis parameters.

Table 2
Estimated power for $T_2$ when $k=2,\ m=2$, and $n_1=n_2=20$

| | $\Omega_1=\Omega_2=I_2$ | $\Omega_1=I_2,\ \Omega_2=V$ | $\Omega_1=I_2,\ \Omega_2=C$ |
|---|---|---|---|
| Multivariate normal | 0.053 | 0.707 | 0.241 |
| Multivariate $t$ | 0.058 | 0.357 | 0.171 |

The fourth nonnormal elliptical case tabulated in Table 1(e) was chosen so as to illustrate that the normal-theory test can sometimes be very conservative. Here again we utilized the stochastic decomposition $x=rTu$ for a random vector $x$ having an elliptical distribution with mean $\mu$ and covariance matrix $\Omega$. We used $T=I_m$ and we chose $r$ as the random variable having density, $f(r)=r^3$, for $0<r<2^{1/2}$, and 0 elsewhere. For this choice of a distribution for $x$, using (4), we find that $\kappa=-0.438$ when $m=2$ and $\kappa=-0.196$ when $m=5$, so in both cases the distribution of $x$ is platykurtic. The results in Table 1(e) show that in both cases $T_1$ yields significance levels that are much too small. This problem is worse when $m=2$ and this is to be expected due to the smaller value of $\kappa$.

The final portion of Table 1 contains some results for a case in which each of the three populations have the same nonelliptical distribution. The particular distribution chosen was the one used by Zhang and Boos (1992). Each component was a $\mathrm{N}(0,1)$ random variable with probability 0.9 and a $\chi_2^2$ random variable with probability 0.1, where the components are independent. From Table 1(f), we see that all of the procedures except the normal-theory procedure perform reasonably well for $m=2$. However, when $m=5$, the procedures based on $T_2$ and $T_3$ are overly conservative, while $T_4$ exhibits its typical slow convergence to the asymptotic distribution.

Table 2 has some additional simulation results for $T_2$ when $k=2$, $m=2$, and $n_1=n_2=20$. The first column gives estimates of the actual type I error probability when sampling from normal distributions and multivariate $t$ distributions with 5 degrees of freedom. The remaining two columns contain power estimates when one of the covariance matrices is the identity matrix and the other is either $V=\mathrm{diag}(2,4)$ or

$$C=\begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}.$$

Upon comparing these results with those given by Zhang and Boos (1992), we find that their bootstrap procedure and $T_2$ have similar power levels when $\Omega_2=C$; when $\Omega_2=V$, $T_2$ is more powerful for normal populations, but less powerful for the multivariate $t$ distribution.

## 4. Conclusions

Of the tests developed in this paper, the one based on $T_2$ appears to be the most useful. It has a computationally simple form, offers a more robust alternative to the normal-theory tests and, in most cases, performs a little better than the test based on

$T_3$. On the other hand, the test based on $T_4$ is not appropriate unless sample sizes are very large or $m$ is very small.

## References

Fang, K.T., Kotz, S., Ng, K.W., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall, New York.

Flury, B., 1988. Common Principal Components and Related Multivariate Models. Wiley, New York.

Hopkins, J.W., Clay, P.P.F., 1963. Some empirical distributions of bivariate $T^2$ and homoscedasticity criteron $M$ under unequal variance and leptokurtosis. J. Amer. Statist. Assoc. 58, 1048–1053.

James, G.S., 1954. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. Biometrika 41, 19–43.

Krzanowski, W.J., 1979. Between-group comparison of principal components. J. Amer. Statist. Assoc. 74, 703–707.

Layard, M.W.J., 1974. A Monte Carlo comparison of tests for equality of covariance matrices. Biometrika 61, 461–465.

Magnus, J.R., 1988. Linear Structures, Oxford University Press, New York.

Moore, D.S., 1977. Generalized inverses, Wald's method, and the construction of chi-squared tests of fit. J. Amer. Statist. Assoc. 72, 131–137.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

Muirhead, R.J., Waternaux, C.M., 1980. Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. Biometrika 67, 31–43.

O'Brien, P.C., 1992. Robust procedures for testing equality of covariance matrices. Biometrics 48, 819–827.

Olson, C.L., 1974. Comparative robustness of six tests in multivariate analysis of variance. J. Amer. Statist. Assoc. 69, 894–908.

Schott, J.R., 1991. Some tests for common principal component subspaces in several groups. Biometrika 78, 771–7.

Tiku, M.L., Balakrishnan, N., 1985. Testing equality of variance-covariance matrices the robust way. Comm. Statist. A 14, 3033–3051.

Tyler, D.E., 1983. Robustness and efficiency properties of scatter matrices. Biometrika 79, 411–420.

Zhang, J., Boos, D.D., 1992. Bootstrap critical values for testing homogeneity of covariance matrices. J. Amer. Statist. Assoc. 87, 425–429.