The statistic of greatest power in a class of test statistics for testing equality of means of two groups without assuming equal covariance matrices

Tetsuto Himeno [1]

Graduate School of Mathematics
Kyushu University
6-10-1 Hakozaki, Higashiku, Fukuoka-city 812-8581, Japan

Abstract

We consider a class of test statistics including the Dempster trace criterion in the case of two groups without assuming equal covariance matrices. The test statistics in the class are valid when the dimension is larger than the sample size. We obtain asymptotic distributions of the test statistics in the class and use these distributions to derive the limiting power in each case. We obtain the most powerful test in the class with respect to this limiting power.

Key words: Dempster trace criterion; Power comparison; Asymptotic distribution; most powerful test

*AMS 2000 subject classification:*Primary 62F05; secondary 62E20

1 Introduction

We consider the model:

$$Y_j = X_j B_j + \varepsilon_j, \quad (j = 1, 2),$$

where Y_j is the $n_j \times p$ observation matrix, X_j is the $n_j \times k$ design matrix of full rank$(X_j) = k$, B_j is the $k \times p$ matrix of regression coefficients, ε_j is the

[1] Fax: +81-92-642-7047.
Email address: `himeno@math.kyushu-u.ac.jp` (Tetsuto Himeno).

$n_j \times p$ error matrix distributed according to $N_{n_j \times p}(O, I_{n_j} \otimes \Sigma_j)$, and ε_1 and ε_2 are mutually independent. We consider the hypothesis

$$H_0 : B_1 = B_2.$$

It is known that MLEs of B_j and Σ_j are given by

$$\hat{B}_j = (X_j' X_j)^{-1} X_j' Y_j \quad \text{and} \quad \hat{\Sigma}_j = \frac{1}{n_j}(Y_j - X\hat{B}_j)'(Y_j - X\hat{B}_j).$$

We note here that if the dimension is larger than the sample size, there are some statistics that are not defined. The Dempster trace criterion is one of the test statistics suitable for this situation. Under the condition of homogeneity of covariance matrices, i.e., $\Sigma_1 = \Sigma_2$, the Dempster trace criterion was proposed as $\mathrm{tr}(S_h)/\mathrm{tr}(S_e)$, where

$$\begin{aligned} S_h &= (\hat{B}_1 - \hat{B}_2)'((X_1'X_1)^{-1} + (X_2'X_2)^{-1})^{-1}(\hat{B}_1 - \hat{B}_2), \\ S_e &= n_1\hat{\Sigma}_1 + n_2\hat{\Sigma}_2. \end{aligned}$$

In this paper, we utilize a more general class of statistics, relaxing the homogeneity constraint. Consider the test statistics

$$\frac{\mathrm{tr}\{(\hat{B}_1 - \hat{B}_2)' A (\hat{B}_1 - \hat{B}_2)\}}{\mathrm{tr}(a n_1 \hat{\Sigma}_1 + b n_2 \hat{\Sigma}_2)},$$

where A is a $k \times k$ symmetric matrix, and a and b are non-negative constants. This class of test statistics includes the Dempster trace criterion under the condition $\Sigma_1 = \Sigma_2$. In this paper, in order to derive the limiting distribution, we assume the conditions

$$\begin{aligned} \mathrm{tr}\{(B_1 - B_2)' A (B_1 - B_2)\} = O(p), \quad \mathrm{tr}(\Sigma_1^i \Sigma_2^j) = O(p), \\ \mathrm{tr}\{A(X_i'X_i)^{-1}\} = O(1), \quad a = O(1), \quad b = O(1), \end{aligned} \tag{1}$$

where $i, j = 0, 1, 2$ and $0 \le i + j \le 2$. We derive the limiting power of this test statistic and find the best parameter A for some values of a and b under the framework:

$$k : \text{fixed}, \;\; N_i \to \infty, \;\; p \to \infty, \;\; \frac{p}{N_i} \to c \in (0, \infty), \quad (i = 1, 2). \tag{2}$$

In Section 2, we obtain the asymptotically most powerful test. In Section 3, we examine the power of the resulting test in a numerical simulation. We introduce some asymptotic results in the appendix.

2 Asymptotically most powerful test statistic

In this section, we derive the limiting power of the test statistic. Using the limiting power, the test statistic of greatest power is obtained.

We define T and T^* as

$$\begin{aligned} T =& \sqrt{p}\left(\frac{n\mathrm{tr}\{(\hat{B}_1-\hat{B}_2)'A(\hat{B}_1-\hat{B}_2)\}}{\mathrm{tr}(an_1\hat{\Sigma}_1+bn_2\hat{\Sigma}_2)}\right.\\ &\left.-\frac{\mathrm{tr}(\Sigma_1)\mathrm{tr}\{A(X_1'X_1)^{-1}\}+\mathrm{tr}(\Sigma_2)\mathrm{tr}\{A(X_2'X_2)^{-1}\}}{\frac{a(N_1-k)}{n}\mathrm{tr}(\Sigma_1)+\frac{b(N_2-k)}{n}\mathrm{tr}(\Sigma_2)}\right),\\ T^* =& T-\sqrt{p}\left(\frac{\mathrm{tr}\{(B_1-B_2)'A(B_1-B_2)\}}{\frac{a(N_1-k)}{n}\mathrm{tr}(\Sigma_1)+\frac{b(N_2-k)}{n}\mathrm{tr}(\Sigma_2)}\right), \end{aligned}$$

where $n = N_1 + N_2$. Let

$$\delta = T - T^* = \sqrt{p}\frac{\mathrm{tr}\{(B_1-B_2)'A(B_1-B_2)\}}{\frac{a(N_1-k)}{n}\mathrm{tr}(\Sigma_1)+\frac{b(N_2-k)}{n}\mathrm{tr}(\Sigma_2)}.$$

Then the power of the test at significance level α when using T can be expressed as

$$P = \Pr\left(\frac{T}{\sigma} > z_\alpha\right) + o(1) = \Pr\left(\frac{T^*}{\sigma_*} > \frac{\sigma z_\alpha - \delta}{\sigma_*}\right) + o(1),$$

where

$$\begin{aligned} \sigma^2 =& \left(2\frac{1}{p}\mathrm{tr}(\Sigma_1^2)\mathrm{tr}\{(A(X_1'X_1)^{-1})^2\}\right.\\ &+4\frac{1}{p}\mathrm{tr}(\Sigma_1\Sigma_2)\mathrm{tr}\{A(X_1'X_1)^{-1}A(X_2'X_2)^{-1}\}\\ &\left.+2\frac{1}{p}\mathrm{tr}(\Sigma_2^2)\mathrm{tr}\{(A(X_2'X_2)^{-1})^2\}\right)\\ &\times\left(\frac{a(N_1-k)}{np}\mathrm{tr}(\Sigma_1)+\frac{b(N_2-k)}{np}\mathrm{tr}(\Sigma_2)\right)^{-2},\\ \sigma_*^2 =& \left(2\frac{1}{p}\mathrm{tr}(\Sigma_1^2)\mathrm{tr}\{(A(X_1'X_1)^{-1})^2\}\right.\\ &+4\frac{1}{p}\mathrm{tr}(\Sigma_1\Sigma_2)\mathrm{tr}\{A(X_1'X_1)^{-1}A(X_2'X_2)^{-1}\} \end{aligned}$$

$$
\begin{aligned}
&+2\frac{1}{p}\mathrm{tr}(\Sigma_2^2)\mathrm{tr}\{(A(X_2'X_2)^{-1})^2\} \\
&+\frac{4}{p}\mathrm{tr}\{(B_1-B_2)'A(X_1'X_1)^{-1}A(B_1-B_2)\Sigma_1\} \\
&+\frac{4}{p}\mathrm{tr}\{(B_1-B_2)'A(X_2'X_2)^{-1}A(B_1-B_2)\Sigma_2\}\Bigg) \\
&\times\left(\frac{a(N_1-k)}{np}\mathrm{tr}(\Sigma_1)+\frac{b(N_2-k)}{np}\mathrm{tr}(\Sigma_2)\right)^{-2},
\end{aligned}
$$

and z_α is the upper 100α % point of the standard normal distribution (See Appendix). If the order of δ is greater than 1, the asymptotic power is 1. We assume that $\mathrm{tr}\{(B_1-B_2)'A(B_1-B_2)\}=O(\sqrt{p})$. In addition, we assume that

$$
\mathrm{tr}\{(B_1-B_2)'A(X_i'X_i)^{-1}A(B_1-B_2)\Sigma_i\}=O(\sqrt{p}),\quad (i=1,2).
$$

Then σ_* converges to σ and

$$
\lim_{n,p\to\infty}P=\Phi\left(\frac{\delta}{\sigma}-z_\alpha\right),
$$

where $\Phi(\cdot)$ is the standard normal distribution function. We want to maximize

$$
\begin{aligned}
\left(\frac{\delta}{\sigma}\right)^2 &=\frac{1}{2}\mathrm{tr}\{(B_1-B_2)(B_1-B_2)'A\}^2\Big(\mathrm{tr}(\Sigma_1^2)\mathrm{tr}\{(A(X_1'X_1)^{-1})^2\} \\
&\quad+2\mathrm{tr}(\Sigma_1\Sigma_2)\mathrm{tr}\{A(X_1'X_1)^{-1}A(X_2'X_2)^{-1}\} \\
&\quad+\mathrm{tr}(\Sigma_2^2)\mathrm{tr}\{(A(X_2'X_2)^{-1})^2\}\Big)^{-1} \\
&=\frac{1}{2}(\mathrm{Vec}(A))'\mathrm{vec}((B_1-B_2)(B_1-B_2)') \\
&\quad\times(\mathrm{vec}((B_1-B_2)(B_1-B_2)'))'\mathrm{Vec}(A) \\
&\quad\times\Big((\mathrm{Vec}(A))'\{\mathrm{tr}(\Sigma_1^2)(X_1'X_1)^{-1}\otimes(X_1'X_1)^{-1} \\
&\quad+\mathrm{tr}(\Sigma_1\Sigma_2)(X_1'X_1)^{-1}\otimes(X_2'X_2)^{-1} \\
&\quad+\mathrm{tr}(\Sigma_1\Sigma_2)(X_2'X_2)^{-1}\otimes(X_1'X_1)^{-1} \\
&\quad+\mathrm{tr}(\Sigma_2^2)(X_2'X_2)^{-1}\otimes(X_2'X_2)^{-1}\}\mathrm{Vec}(A)\Big)^{-1},
\end{aligned}
$$

where $\mathrm{Vec}(\cdot)$ is the Vec operator which stacks the columns of a matrix and $\otimes$ is the Kronecker product. This expression does not depend on a and b. Let $a=b=1$ for simplicity. To maximize $(\delta/\sigma)^2$, $\mathrm{Vec}(A)$ is the eigenvector corresponding to the maximum eigenvalue of

$$
\{\mathrm{tr}(\Sigma_1^2)(X_1'X_1)^{-1}\otimes(X_1'X_1)^{-1}+\mathrm{tr}(\Sigma_1\Sigma_2)(X_1'X_1)^{-1}\otimes(X_2'X_2)^{-1}
$$

$$
\begin{aligned}
&+\mathrm{tr}(\Sigma_1\Sigma_2)(X_1'X_1)^{-1}\otimes(X_2'X_2)^{-1}+\mathrm{tr}(\Sigma_2^2)(X_2'X_2)^{-1}\otimes(X_2'X_2)^{-1}\}^{-1}\\
&\times\mathrm{vec}((B_1-B_2)(B_1-B_2)')(\mathrm{vec}((B_1-B_2)(B_1-B_2)'))'.
\end{aligned}
$$

The rank of this matrix is 1. We obtain

$$
\begin{aligned}
\mathrm{Vec}(A)=c\{&\mathrm{tr}(\Sigma_1^2)(X_1'X_1)^{-1}\otimes(X_1'X_1)^{-1}\\
&+\mathrm{tr}(\Sigma_1\Sigma_2)(X_1'X_1)^{-1}\otimes(X_2'X_2)^{-1}\\
&+\mathrm{tr}(\Sigma_1\Sigma_2)(X_2'X_2)^{-1}\otimes(X_1'X_1)^{-1}\\
&+\mathrm{tr}(\Sigma_2^2)(X_2'X_2)^{-1}\otimes(X_2'X_2)^{-1}\}^{-1}\\
&\times\mathrm{vec}((B_1-B_2)(B_1-B_2)'),
\end{aligned}
$$

where c is a positive number such that

$$
\mathrm{tr}\{A(X_i'X_i)^{-1}\}=O(1),\quad (i=1,2).
$$

Hence we obtain the following theorem.

Theorem 2.1. *Let $A=c(a_{ij})$, where c is a positive number. We define the statistic of A by*

$$
\begin{aligned}
\hat{a}_{ij}=(\boldsymbol{e}_j\otimes\boldsymbol{e}_i)'\{&\widehat{\mathrm{tr}(\Sigma_1^2)}(X_1'X_1)^{-1}\otimes(X_1'X_1)^{-1}\\
&+\mathrm{tr}(\hat{\Sigma}_1\hat{\Sigma}_2)(X_1'X_1)^{-1}\otimes(X_2'X_2)^{-1}\\
&+\mathrm{tr}(\hat{\Sigma}_1\hat{\Sigma}_2)(X_2'X_2)^{-1}\otimes(X_1'X_1)^{-1}\\
&+\widehat{\mathrm{tr}(\Sigma_2^2)}(X_2'X_2)^{-1}\otimes(X_2'X_2)^{-1}\}^{-1}\\
&\times\mathrm{vec}((\hat{B}_1-\hat{B}_2)(\hat{B}_1-\hat{B}_2)'),
\end{aligned}
$$

where $\boldsymbol{e}_i$ is the $k\times 1$ vector

$$
\boldsymbol{e}_i=(0,\cdots,0,1,0,\cdots,0)'
$$

with the 1 in the ith place and

$$
\widehat{\mathrm{tr}(\Sigma_i^2)}=\mathrm{tr}(\hat{\Sigma}_i^2)-\{\mathrm{tr}(\hat{\Sigma}_i)\}^2/(n_i-k),\quad (i=1,2).
$$

Then

$$
\frac{\mathrm{tr}\{(\hat{B}_1-\hat{B}_2)'\hat{A}(\hat{B}_1-\hat{B}_2)\}}{\mathrm{tr}(n_1\hat{\Sigma}_1+n_2\hat{\Sigma}_2)},\tag{3}
$$

is asymptotically the most powerful test in this class.

If d is a positive number and $\Sigma_1=d\Sigma_2$, then we obtain

$$\begin{aligned}\mathrm{Vec}(A) &= c\{(d(X_1'X_1)^{-1} + (X_2'X_2)^{-1}) \otimes (d(X_1'X_1)^{-1} + (X_2'X_2)^{-1})\}^{-1} \\ &\quad \times \mathrm{vec}((B_1 - B_2)(B_1 - B_2)') \\ &= c(d(X_1'X_1)^{-1} + (X_2'X_2)^{-1})^{-1} \otimes (d(X_1'X_1)^{-1} + (X_2'X_2)^{-1})^{-1} \\ &\quad \times \mathrm{vec}((B_1 - B_2)(B_1 - B_2)').\end{aligned}$$

This expression may be written as

$$\begin{aligned}A &= c(d(X_1'X_1)^{-1} + (X_2'X_2)^{-1})^{-1}(B_1 - B_2)(B_1 - B_2)' \\ &\quad \times (d(X_1'X_1)^{-1} + (X_2'X_2)^{-1})^{-1}.\end{aligned}$$

Therefore, using the estimate of A:

$$\begin{aligned}\hat{A} &= c(\hat{d}(X_1'X_1)^{-1} + (X_2'X_2)^{-1})^{-1}(\hat{B}_1 - \hat{B}_2)(\hat{B}_1 - \hat{B}_2)' \\ &\quad \times (\hat{d}(X_1'X_1)^{-1} + (X_2'X_2)^{-1})^{-1},\end{aligned}$$

where $\hat{d} = \mathrm{tr}(\hat{\Sigma}_2^{-1}\hat{\Sigma}_1)/p$, we expect that the power of

$$\frac{\mathrm{tr}\{(\hat{B}_1 - \hat{B}_2)'\hat{A}(\hat{B}_1 - \hat{B}_2)\}}{\mathrm{tr}(n_1\hat{\Sigma}_1 + n_2\hat{\Sigma}_2)} = \frac{\mathrm{tr}(S_h^2)}{\mathrm{tr}(S_e)}$$

is large. In particular, under the condition of homogeneity of covariance matrices, it is the case that $d = 1$. When $X_1 = X_2$, we derive a similar result. Therefore, we obtain the following corollary.

Corollary 2.2. *If $\Sigma_1 = \Sigma_2$ or $X_1 = X_2$, then*

$$\frac{\mathrm{tr}\{(\hat{B}_1 - \hat{B}_2)'\hat{A}(\hat{B}_1 - \hat{B}_2)\}}{\mathrm{tr}(n_1\hat{\Sigma}_1 + n_2\hat{\Sigma}_2)} = \frac{\mathrm{tr}(S_h^2)}{\mathrm{tr}(S_e)} \tag{4}$$

is asymptotically the most powerful test in this class.

3 Numerical simulation

In this section, we examine the optimality of the test statistic (3) and (4).

We compare the power of the test statistics resulting from (3), (4) and $\mathrm{tr}(S_h)/\mathrm{tr}(S_e)$ numerically. Under the null hypothesis, we derive the upper 5 % point by Monte Carlo simulation. Using the upper 5 % point, we derive the

power under the non-null hypothesis numerically.
We chose the value of the parameters as follow:

$$
\begin{aligned}
&N_1 = N_2(= n),\\
&(k, n, p) = (2, 20, 20), (2, 20, 40), (2, 40, 20), (2, 40, 40),\\
&(X_1, X_2) = (I_k \otimes \mathbf{1}_{n/k}, I_k \otimes \mathbf{1}_{n/k}), \left(\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \otimes \mathbf{1}_{n/k}, I_k \otimes \mathbf{1}_{n/k} \right),\\
&(\Sigma_1, \Sigma_2) = (I_p, I_p), (I_p, 1.5 I_p),\\
&B_2 - B_1 = 1.2(\mathbf{1}_k, \cdots, \mathbf{1}_k),
\end{aligned}
$$

where $\mathbf{1}_m$ is the $m \times 1$ vector with all elements equal to 1. In Table (4.1), we assume that $(X_1, X_2) = (I_k \otimes \mathbf{1}_{n/k}, I_k \otimes \mathbf{1}_{n/k})$. In Table (4.2), we assume that

$$
(X_1, X_2) = \left(\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \otimes \mathbf{1}_{n/k}, I_k \otimes \mathbf{1}_{n/k} \right)
$$

. The power of the test statistic (4) is greater than that of the other test statistics. We see that the gap of the covariance matrices and their statistics causes the reduction of the power of the test statistic (3). So, even if $\Sigma_1 \neq \Sigma_2$ and $X_1 \neq X_2$, the power of the test statistic (4) is greatest.

Acknowledgements
The author is deeply grateful to Prof. Y. Fujikoshi, Chuo University; Prof. H. Wakaki and Dr. H. Yanagihara, Hiroshima University, for their valuable comments.

Appendix
We derive the limiting distribution of the test statistics under an assumption and the framework (2).
Let U and V be defined by

$$
\begin{aligned}
U &= \frac{1}{\sqrt{p}}(\mathrm{tr}\{(\hat{B}_1 - \hat{B}_2)' A (\hat{B}_1 - \hat{B}_2)\} - \mathrm{tr}\{(B_1 - B_2)' A (B_1 - B_2)\}\\
&\qquad - \mathrm{tr}(\Sigma_1)\mathrm{tr}\{A(X_1'X_1)^{-1}\} - \mathrm{tr}(\Sigma_2)\mathrm{tr}\{A(X_2'X_2)^{-1})\},\\
V &= \frac{1}{\sqrt{np}}(\mathrm{tr}(a n_1 \hat{\Sigma}_1 + b n_2 \hat{\Sigma}_2) - a(N_1 - k)\mathrm{tr}(\Sigma_1) - b(N_2 - k)\mathrm{tr}(\Sigma_2)).
\end{aligned}
$$

Then U and V are asymptotically normal, and T^* is expressed by

Table 1

$(X_1, X_2) = (I_k \otimes \mathbf{1}_{n/k}, I_k \otimes \mathbf{1}_{n/k})$

Σ_2	n	p	$\mathrm{tr}(S_h)/\mathrm{tr}(S_e)$	(3)	(4)
I_p	20	20	0.212	0.214	0.231
I_p	20	40	0.316	0.308	0.344
I_p	40	20	0.459	0.484	0.504
I_p	40	40	0.701	0.738	0.762
$1.5I_p$	20	20	0.175	0.171	0.185
$1.5I_p$	20	40	0.251	0.251	0.272
$1.5I_p$	40	20	0.362	0.371	0.391
$1.5I_p$	40	40	0.560	0.587	0.611

Table 2

$(X_1, X_2) = \big(((1,1)', (0,1)) \otimes \mathbf{1}_{n/k}, I_k \otimes \mathbf{1}_{n/k}\big)$

Σ_2	n	p	$\mathrm{tr}(S_h)/\mathrm{tr}(S_e)$	(3)	(4)
I_p	20	20	0.373	0.376	0.414
I_p	20	40	0.561	0.577	0.616
I_p	40	20	0.758	0.787	0.810
I_p	40	40	0.944	0.960	0.969
$1.5I_p$	20	20	0.270	0.243	0.304
$1.5I_p$	20	40	0.404	0.381	0.475
$1.5I_p$	40	20	0.577	0.561	0.642
$1.5I_p$	40	40	0.825	0.823	0.878

$$
\begin{aligned}
T^* = n\sqrt{p} \Bigg\{ & \Big(\sqrt{p} U + \mathrm{tr}\{(B_1 - B_2)' A (B_1 - B_2)\} \\
& + \mathrm{tr}(\Sigma_1)\mathrm{tr}\{A(X_1'X_1)^{-1}\} + \mathrm{tr}(\Sigma_2)\mathrm{tr}\{A(X_2'X_2)^{-1}\} \Big) \\
& \times \Big(\sqrt{np} V + a(N_1 - k)\mathrm{tr}(\Sigma_1) + b(N_2 - k)\mathrm{tr}(\Sigma_2) \Big)^{-1} \\
& - \Big(\mathrm{tr}(\Sigma_1)\mathrm{tr}\{A(X_1'X_1)^{-1}\} + \mathrm{tr}(\Sigma_2)\mathrm{tr}\{A(X_2'X_2)^{-1}\} \\
& + \mathrm{tr}\{(B_1 - B_2)' A (B_1 - B_2)\} \Big) \\
& \times \Big(a(N_1 - k)\mathrm{tr}(\Sigma_1) + b(N_2 - k)\mathrm{tr}(\Sigma_2) \Big)^{-1} \Bigg\} \\
= & \frac{U}{\frac{a(N_1-k)}{np}\mathrm{tr}(\Sigma_1) + \frac{b(N_2-k)}{np}\mathrm{tr}(\Sigma_2)} + o_p(1).
\end{aligned}
$$

We obtain the following theorem.

Theorem A.1. *Under the framework (2) and assuming the condition (1), it follows that*

$$\frac{T}{\sigma} \xrightarrow{d} N(0,1)$$

under the null hypothesis and

$$\frac{T^*}{\sigma_*} \xrightarrow{d} N(0,1)$$

under the non-null hypothesis, where $\xrightarrow{d}$ *denotes convergence in distribution and*

$$\begin{aligned}
\sigma^2 = &\left(2\frac{1}{p}\mathrm{tr}(\Sigma_1^2)\mathrm{tr}\{(A(X_1'X_1)^{-1})^2\} \right. \\
&+4\frac{1}{p}\mathrm{tr}(\Sigma_1\Sigma_2)\mathrm{tr}\{A(X_1'X_1)^{-1}A(X_2'X_2)^{-1}\} \\
&\left. +2\frac{1}{p}\mathrm{tr}(\Sigma_2^2)\mathrm{tr}\{(A(X_2'X_2)^{-1})^2\} \right) \\
&\times \left(\frac{a(N_1-k)}{np}\mathrm{tr}(\Sigma_1) + \frac{b(N_2-k)}{np}\mathrm{tr}(\Sigma_2) \right)^{-2}, \\
\sigma_*^2 = &\left(2\frac{1}{p}\mathrm{tr}(\Sigma_1^2)\mathrm{tr}\{(A(X_1'X_1)^{-1})^2\} \right. \\
&+4\frac{1}{p}\mathrm{tr}(\Sigma_1\Sigma_2)\mathrm{tr}\{A(X_1'X_1)^{-1}A(X_2'X_2)^{-1}\} \\
&+2\frac{1}{p}\mathrm{tr}(\Sigma_2^2)\mathrm{tr}\{(A(X_2'X_2)^{-1})^2\} \\
&+\frac{4}{p}\mathrm{tr}\{(B_1-B_2)'A(X_1'X_1)^{-1}A(B_1-B_2)\Sigma_1\} \\
&\left. +\frac{4}{p}\mathrm{tr}\{(B_1-B_2)'A(X_2'X_2)^{-1}A(B_1-B_2)\Sigma_2\} \right) \\
&\times \left(\frac{a(N_1-k)}{np}\mathrm{tr}(\Sigma_1) + \frac{b(N_2-k)}{np}\mathrm{tr}(\Sigma_2) \right)^{-2},
\end{aligned}$$

References

[1] A. P. Dempster, A high dimensional two sample significance test, Ann. Math. Statist, 29 (1958), 995-1010.

[2] A. P. Dempster, A significance test for the separation of two highly multivariate small samples, Biometrics, 16 (1960), 41-50.