Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Statistics & Probability Letters 66 (2004) 161–169
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem

K. Krishnamoorthy*, Jianqi Yu

Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504-1010, USA

Received June 2003

Abstract

A new test to the multivariate Behrens–Fisher problem is obtained by modifying Nel and Van der Merwe's (Comm. Statist. Theory Methods 15 (1986) 3719) test. The new test is affine invariant and it simplifies to the Welch's approximate solution to the univariate case. The merits of the new test and two existing invariant tests are evaluated using Monte Carlo method. Monte Carlo comparison shows that the new test is as powerful as the other two methods while controlling the sizes satisfactorily.

Keywords: Welch approximation; Multiple comparison; Size; Power

1. Introduction

Let $X_{i1},\ldots,X_{iN_i}$ be a sample of vector observations from a p-variate normal population with mean vector $\boldsymbol{\mu}_i$ and covariance matrix Σ_i, $\mathrm{N}_p(\boldsymbol{\mu}_i,\Sigma_i)$, $i=1,2$. The problem of testing equality of the mean vectors without assuming equality of the covariance matrices is referred to as the Behrens–Fisher problem. The hypotheses of interest are

$$\mathrm{H}_0:\boldsymbol{\mu}_1=\boldsymbol{\mu}_2 \text{ vs. } \mathrm{H}_a:\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2. \tag{1}$$

Define

$$\bar{X}_i=\frac{1}{N_i}\sum_{j=1}^{N_i}X_{ij} \quad \text{and} \quad S_i=\frac{1}{n_i}\sum_{j=1}^{N_i}(X_{ij}-\bar{X}_i)(X_{ij}-\bar{X}_i)', \quad i=1,2,$$

* Corresponding author. Tel.: +1-337-4825-283; fax: +1-337-4825-346.
E-mail address: krishna@louisiana.edu (K. Krishnamoorthy).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2003.10.012

where $n_i = N_i - 1$, $i = 1,2$. Let $\tilde{S}_i = S_i/N_i$ and $\tilde{\Sigma}_i = \Sigma_i/N_i$, $i = 1,2$. We note that

$$\bar{X}_i \sim \mathrm{N}_p(\boldsymbol{\mu}_i, \tilde{\Sigma}_i) \quad \text{and} \quad \tilde{S}_i \sim W_p(n_i, \tilde{\Sigma}_i/n_i),\ i = 1,2,$$

and they are all independent. We also assume that $n_i \geqslant p$ so that S_i^{-1} exists, $i=1,2$, with probability one. Because the above distributions are affine invariant and $\mathrm{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ is equivalent to $\mathrm{H}_a : A\boldsymbol{\mu}_1 = A\boldsymbol{\mu}_2$ for any nonsingular matrix, a practical testing method should be nonsingular invariant; otherwise, for a given data set, the p-value for testing $\mathrm{H}_{01} : A_1\boldsymbol{\mu}_1 = A_1\boldsymbol{\mu}_2$ may be different from the one for testing $\mathrm{H}_{02} : A_2\boldsymbol{\mu}_1 = A_2\boldsymbol{\mu}_2$, when A_1 and A_2 are different nonsingular matrices. As a consequence, the conclusions could be different even though H_{01} and H_{02} are equivalent. Furthermore, other inferential procedures, such as multiple comparisons, can be readily obtained from an invariant test. In addition, as shown in Section 3, nonsingular transformation reduces the parameter space considerably so that the merit of an invariant solution can be evaluated numerically over a wide range of parameter space.

We shall now describe a typical approximate approach based on the popular Welch's (1947) "approximate degrees of freedom solution" to the univariate case. Let $\tilde{S} = \tilde{S}_1 + \tilde{S}_2$. A natural invariant test statistic based on the union–intersection principle is

$$T^2 = \max_{\mathbf{a} \in \mathbb{R}^p} \frac{(\mathbf{a}'(\bar{X}_1 - \bar{X}_2))^2}{\mathbf{a}'(\tilde{S}_1 + \tilde{S}_2)\mathbf{a}} = (\bar{X}_1 - \bar{X}_2)'\tilde{S}^{-1}(\bar{X}_1 - \bar{X}_2). \tag{2}$$

Since the above T^2 is similar to the one for the one-sample case, many authors approximated its null distribution by a constant times F distribution with numerator degrees of freedom p and the approximate denominator degrees of freedom depending on $(n_1, n_2, \tilde{S}_1, \tilde{S}_2)$; in Yao's (1965) approach the degrees of freedom are also dependent on $\bar{X}_1 - \bar{X}_2$. Therefore, such an approximate solution based on T^2 is invariant only when the denominator degree of freedom is also nonsingular invariant. Among the approximate solutions based on T^2, the solutions due to James (1954), Yao (1965) and Johansen (1980) are invariant, whereas the solution due to Nel and Van der Merwe (1986) is not invariant (see Section 2). Kim's (1992) approximate solution is not based on T^2, and is obtained by combining the ellipsoids based on the individual samples. Kim's test statistic is given by

$$(\bar{X}_1 - \bar{X}_2)'(\tilde{S}_1 + r^2\tilde{S}_2 + 2rQ(\tilde{S}_1, \tilde{S}_2))^{-1}(\bar{X}_1 - \bar{X}_2),$$

where $r = (\det(\tilde{S}_1\tilde{S}_2^{-1}))^{1/(2p)}$, $Q(\tilde{S}_1, \tilde{S}_2) = \tilde{S}_2^{1/2}(\tilde{S}_2^{-1/2}\tilde{S}_1\tilde{S}_2^{-1/2})^{1/2}\tilde{S}_2^{1/2}$, and $\tilde{S}_2^{1/2}\tilde{S}_2^{1/2} = \tilde{S}_2$. Although Kim (1992, p.174) claims that his procedure is affine invariant, it can be easily verified that $Q(A\tilde{S}_1A', A\tilde{S}_2A')$ cannot be written as $AQ(\tilde{S}_1, \tilde{S}_2)A'$ for an arbitrary nonsingular matrix A, and as a consequence, Kim's test statistic is not invariant.

Regarding comparison of these approximate solutions, simulation studies by Subramaniam and Subramaniam (1973) have indicated that Bennett's (1951) test has poor power for unequal sample sizes even though it is exact. The size for Yao's (1965) test is closer to the nominal level than is the size for James' (1954) test. Christensen and Rencher (1997) compared several methods including some that are not invariant. Their simulation studies indicated that Kim and Nel and Van der Merwe's procedures had the highest power among the procedures whose sizes were not inflated. However, as we already noted these procedures are not invariant, and the conclusions may not be valid over the entire parameter space $\{(\Sigma_1,\Sigma_2):\Sigma_1>0,\Sigma_2>0\}$. Nel et al. (1990) derived the exact[1] null distribution of T^2, but because it is computationally intractable, it is of no use for practical applications.

The purpose of this note is to identify a simple invariant solution that is satisfactory (with respect to size and power) for practical purpose. In the following section, we describe invariant methods due to Yao (1965), Johansen (1980) and a new method. The new method is obtained by modifying Nel and Van der Merwe's (1986) approximate degrees of freedom solution so that the resulting method is invariant. We refer to this approach as the modified Nel and Van der Merwe (MNV) approach. The methods are compared using Monte Carlo simulation consisting of 100,000 runs in Section 3. Our comparison studies indicate that the sizes for the MNV test are closer to the nominal level than are the sizes of other methods, and it is as powerful as other methods. Some concluding remarks are given in Section 4.

2. Invariant tests

In the following we describe three invariant solutions using the notations given in the introduction.

2.1. Yao's (1965)

This method is a multivariate extension of the Welch "approximate degrees of freedom" solution. The approximate degrees of freedom is given by

$$v=\left[\frac{1}{n_1}\left(\frac{\bar{X}_d'\tilde{S}^{-1}\tilde{S}_1\tilde{S}^{-1}\bar{X}_d}{\bar{X}_d'\tilde{S}^{-1}\bar{X}_d}\right)^2+\frac{1}{n_2}\left(\frac{\bar{X}_d'\tilde{S}^{-1}\tilde{S}_2\tilde{S}^{-1}\bar{X}_d}{\bar{X}_d'\tilde{S}^{-1}\bar{X}_d}\right)^2\right]^{-1},$$

where $\bar{X}_d=\bar{X}_1-\bar{X}_2$ and $\tilde{S}=\tilde{S}_1+\tilde{S}_2$. Based on this approximate degrees of freedom, we have $T^2\sim vpF_{p,v-p+1}/(v-p+1)$ approximately.

2.2. Johansen's (1980)

We shall describe Johansen's test as given in Tang and Algina (1993). The test statistic T^2 is approximated by $qF_{p,v}$, where $q=p+2D-6D/[p(p-1)+2]$, $v=p(p+2)/(3D)$,

$$D=\frac{1}{2}\sum_{i=1}^{2}\{\mathrm{tr}[(I-(\tilde{S}_1^{-1}+\tilde{S}_2^{-1})^{-1}\tilde{S}_i^{-1})^2]+[\mathrm{tr}(I-(\tilde{S}_1^{-1}+\tilde{S}_2^{-1})^{-1}\tilde{S}_i^{-1})]^2\}/n_i$$

and $F_{m,n}$ denotes the F distribution with degrees of freedom m and n.

[1] The distribution function is based on an estimate of the exact pdf of T^2.

2.3. The MNV method

This method is based on the approximation to the distribution of $\tilde{S}_1 + \tilde{S}_2$ given in Nel and Van der Merwe (1986). Recall that $\tilde{S}_1 \sim W_p(n_1, \tilde{\Sigma}_1/n_1)$ independently of $\tilde{S}_2 \sim W_p(n_2, \tilde{\Sigma}_2/n_2)$. Nel and Van der Merwe showed that

$$\tilde{S} = \tilde{S}_1 + \tilde{S}_2 \sim W_p\left(f, \frac{1}{f}\tilde{\Sigma}\right) \text{ approximately,} \tag{3}$$

where $\tilde{\Sigma} = \tilde{\Sigma}_1 + \tilde{\Sigma}_2$ and

$$f = \frac{\mathrm{tr}(\tilde{\Sigma}^2) + [\mathrm{tr}(\tilde{\Sigma})]^2}{(1/n_1)\{\mathrm{tr}(\tilde{\Sigma}_1^2) + [\mathrm{tr}(\tilde{\Sigma}_1)]^2\} + (1/n_2)\{\mathrm{tr}(\tilde{\Sigma}_2^2) + [\mathrm{tr}(\tilde{\Sigma}_2)]^2\}}. \tag{4}$$

Notice that the T^2 statistic in (2) is equivalent to $(\bar{X}_1 - \bar{X}_2)'\tilde{\Sigma}^{-1/2}(\tilde{\Sigma}^{-1/2}\tilde{S}\tilde{\Sigma}^{-1/2})^{-1}\tilde{\Sigma}^{-1/2}(\bar{X}_1 - \bar{X}_2)$ and, under H$_0$, $\tilde{\Sigma}^{-1/2}(\bar{X}_1 - \bar{X}_2) \sim N_p(0, I_p)$. Thus, if we show that $\tilde{\Sigma}^{-1/2}\tilde{S}\tilde{\Sigma}^{-1/2}$ follows a Wishart distribution approximately, then the usual argument for deriving the distribution of one-sample T^2 statistic (e.g. see Muirhead, 1982, p.98) can be used to find an approximation to the distribution of T^2 in (2). To find an approximation to the distribution of $\tilde{\Sigma}^{-1/2}\tilde{S}\tilde{\Sigma}^{-1/2}$, let $\Delta_1 = \tilde{\Sigma}^{-1/2}\tilde{\Sigma}_1\tilde{\Sigma}^{-1/2}$. Then $\tilde{\Sigma}^{-1/2}\tilde{S}_1\tilde{\Sigma}^{-1/2} \sim W_p(n_1, \Delta_1/n_1)$ independently of $\tilde{\Sigma}^{-1/2}\tilde{S}_2\tilde{\Sigma}^{-1/2} \sim W_p(n_2, (I_p - \Delta_1)/n_2)$. Note that under this transformation $\tilde{\Sigma}_1 \to \Delta_1$, $\tilde{\Sigma}_2 \to I_p - \Delta_1$ and $\tilde{\Sigma} \to I_p$. Therefore, it follows from (3) and (4) that

$$V = \tilde{\Sigma}^{-1/2}\tilde{S}_1\tilde{\Sigma}^{-1/2} + \tilde{\Sigma}^{-1/2}\tilde{S}_2\tilde{\Sigma}^{-1/2} = \tilde{\Sigma}^{-1/2}\tilde{S}\tilde{\Sigma}^{-1/2} \sim W_p\left(f^*, \frac{1}{f^*}I_p\right) \quad \text{approximately,} \tag{5}$$

where

$$f^* = \frac{\mathrm{tr}(I_p^2) + [\mathrm{tr}(I_p)]^2}{(1/n_1)\{\mathrm{tr}(\Delta_1^2) + [\mathrm{tr}(\Delta_1)]^2\} + (1/n_2)\{\mathrm{tr}((I_p - \Delta_1)^2) + [\mathrm{tr}(I_p - \Delta_1)]^2\}}. \tag{6}$$

Replacing Δ_1 in (6) by its estimate $\tilde{S}^{-1/2}\tilde{S}_1\tilde{S}^{-1/2}$ and then using the result that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, we get the approximate degrees of freedom as

$$v = \frac{p + p^2}{(1/n_1)\{\mathrm{tr}[(\tilde{S}_1\tilde{S}^{-1})^2] + [\mathrm{tr}(\tilde{S}_1\tilde{S}^{-1})]^2\} + (1/n_2)\{\mathrm{tr}[(\tilde{S}_2\tilde{S}^{-1})^2] + [\mathrm{tr}(\tilde{S}_2\tilde{S}^{-1})]^2\}}, \tag{7}$$

which is nonsingular invariant, and as shown in the appendix $\min\{n_1, n_2\} \leqslant v \leqslant n_1 + n_2$. Thus, V in (5) follows $W_p(v, (1/v)I_p)$ approximately, which implies that T^2 in (2) follows $vpF_{p,v-p+1}/(v-p+1)$ approximately. Nel and Van der Merwe's (1986) solution to the Behrens–Fisher problem involves the degrees of freedom

$$\frac{\mathrm{tr}(\tilde{S}^2) + [\mathrm{tr}(\tilde{S})^2]}{(1/n_1)\{\mathrm{tr}(\tilde{S}_1^2) + [\mathrm{tr}(\tilde{S}_1)]^2\} + (1/n_2)\{\mathrm{tr}(\tilde{S}_2^2) + [\mathrm{tr}(\tilde{S}_2)]^2\}}, \tag{8}$$

which is obtained by replacing the unknown parameters in (4) by their estimates. It can be readily verified that the degrees of freedoms in (7) and (8) are the same when $p = 1$, and hence the MNV test and Nel and Van der Merwe's (1986) test are the same. However, they are different when

$p \geqslant 2$, and expression (8) is not invariant under nonsingular transformation. Therefore, Nel and Van der Merwe (1986) test is not invariant when $p \geqslant 2$. It appears that Nel and Van der Merwe overlooked the fact that, under the transformation in (5), the degrees of freedom $f \to f^*$ and hence the appropriate degrees of freedom should be the v in (7). Furthermore, it is not clear that the degrees of freedom in (8) lies in $[\min\{n_1, n_2\}, n_1 + n_2]$ for all $\tilde{S}_1$, $\tilde{S}_2$ and $n_1, n_2 \geqslant p$. The degrees of freedom must be at least $\min\{n_1, n_2\}$; otherwise the approximate denominator degrees of freedom of the F distribution could be negative.

Remark 1. It should be noted that among the invariant solutions considered above, only Yao's and the MNV solutions reduce to the Welch's approximate degrees of freedom solution for the univariate case. Furthermore, the approximate degrees of freedoms for both solutions fall in $[\min\{n_1, n_2\}, n_1 + n_2]$ for any $p \geqslant 1$.

3. Monte Carlo studies and discussion

The sizes and powers of the three invariant solutions given in Section 2 are estimated using Monte Carlo method. We used IMSL subroutine RNNOA for generating normal random variables, and Fortran subroutine due to Smith and Hocking (1972) to generate Wishart matrices. Each simulation consists of 100,000 runs. Because we are estimating probabilities, the maximum error of an estimate is given by $\pm 2\sqrt{0.5 \times 0.5/100,000} = \pm 0.003$. As pointed out by Yao (1965), there exists a nonsingular matrix M such that $\tilde{\Sigma}_1 = M\Lambda M'$, $\tilde{\Sigma}_2 = M(I - \Lambda)M'$ and $\tilde{\Sigma}_1 + \tilde{\Sigma}_2 = MM'$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $0 < \lambda_1 \leqslant \cdots \leqslant \lambda_p < 1$, and λ_i's are the eigenvalues of $(\tilde{\Sigma}_1 + \tilde{\Sigma}_2)^{-1/2}\tilde{\Sigma}_1(\tilde{\Sigma}_1 + \tilde{\Sigma}_2)^{-1/2}$. Because all three solutions are nonsingular invariants, without loss of generality, we can assume $\Sigma_1 = \Lambda$ and $\Sigma_2 = I - \Lambda$ for comparison purpose. The sizes and powers are computed for $p = 1, 2$, and 10, and noncentrality parameter $\delta = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'(\tilde{\Sigma}_1 + \tilde{\Sigma}_2)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. The mean vectors are chosen such that $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = (\sqrt{\delta/p})\mathbf{1}$, where $\mathbf{1}$ denotes the vector of ones. The estimates of the sizes and powers are presented in Table 1. We observe the following from the estimated sizes and powers:

(1) For the univariate case, the MNV and Yao's tests simplify to the Welch's test, and hence the sizes and powers are computed only for Johansen's test and the MNV test. It is clear that only the MNV test (Welch's test) controls the sizes satisfactorily. The sizes of Johansen's test exceed the nominal level 0.05 considerably.

(2) For $p \geqslant 2$, the sizes of Yao's test are in general inflated and it is clearly inferior to other two tests in controlling Type I error rates.

(3) Between Johansen's test and the MNV test, the sizes of the latter are closer to the nominal level than are the sizes of the former. This is exemplified when all the eigenvalues are equal to 0.9, and smaller of the sample sizes is associated with the "larger" of the covariance matrices. In general, Johansen's test has inflated Type I error rates for higher dimension (see table values when $p = 10$).

(4) The powers of Johansen's test and the MNV test are very close when their sizes are close to the nominal level. See $p = 2$, $(\lambda_1, \lambda_2) = (0.2, 0.5), (0.2, 0.7)$ and $(0.5, 0.5)$; $p = 10, (\lambda_1, \ldots, \lambda_{10}) = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.9)$ and $n_1 = 19, n_2 = 29$. In other situations, because of its inflated sizes, Johansen's test appears to be more powerful than the MNV test.

Table 1
Monte carlo estimates of the sizes and powers of the tests

$n_1=4, n_2=11$							$n_1=4, n_2=4$				
		δ					δ				
λ_1	Test[a]	0	2	4	8	16	0	2	4	8	16
$p=1,\ \alpha=0.05$											
0.1	2	0.059	0.282	0.486	0.768	0.965	0.071	0.272	0.445	0.698	0.923
	3	0.051	0.257	0.455	0.741	0.957	0.054	0.218	0.370	0.615	0.874
0.2	2	0.058	0.279	0.489	0.773	0.968	0.067	0.270	0.452	0.714	0.933
	3	0.050	0.256	0.458	0.748	0.961	0.050	0.221	0.386	0.642	0.896
0.3	2	0.057	0.278	0.486	0.770	0.967	0.061	0.266	0.454	0.722	0.943
	3	0.048	0.254	0.456	0.745	0.960	0.047	0.221	0.392	0.659	0.913
0.4	2	0.055	0.277	0.483	0.768	0.963	0.061	0.265	0.455	0.726	0.946
	3	0.046	0.251	0.451	0.740	0.954	0.046	0.220	0.396	0.668	0.922
0.5	2	0.059	0.278	0.483	0.761	0.961	0.059	0.266	0.453	0.730	0.949
	3	0.049	0.250	0.445	0.727	0.949	0.044	0.221	0.394	0.672	0.926
0.6	2	0.060	0.279	0.474	0.752	0.955	0.060	0.263	0.453	0.727	0.949
	3	0.050	0.248	0.432	0.710	0.937	0.045	0.218	0.394	0.669	0.923
0.7	2	0.064	0.277	0.466	0.736	0.946	0.061	0.267	0.457	0.725	0.942
	3	0.053	0.242	0.417	0.685	0.921	0.046	0.220	0.393	0.661	0.913
0.8	2	0.060	0.255	0.432	0.695	0.937	0.065	0.271	0.455	0.712	0.934
	3	0.054	0.233	0.400	0.657	0.903	0.049	0.222	0.388	0.643	0.897
0.9	2	0.071	0.274	0.448	0.702	0.925	0.070	0.270	0.445	0.698	0.924
	3	0.056	0.224	0.377	0.622	0.877	0.053	0.218	0.370	0.617	0.876
0.95	2	0.074	0.270	0.437	0.689	0.916	0.073	0.269	0.436	0.689	0.918
	3	0.055	0.213	0.358	0.599	0.861	0.053	0.210	0.356	0.598	0.862

$n_1=6, n_2=12$							$n_1=12, n_2=12$				
		δ					δ				
(λ_1, λ_2)	Test[b]	0	2	4	8	16	0	2	4	8	16
$p=2$											
(0.1,0.1)	1	0.051	0.191	0.341	0.612	0.900	0.053	0.193	0.347	0.617	0.900
	2	0.051	0.193	0.345	0.617	0.903	0.052	0.194	0.350	0.621	0.904
	3	0.050	0.190	0.340	0.612	0.901	0.051	0.191	0.346	0.617	0.901
(0.2,0.5)	1	0.049	0.186	0.342	0.623	0.907	0.048	0.196	0.360	0.646	0.920
	2	0.051	0.190	0.347	0.628	0.910	0.049	0.197	0.362	0.648	0.921
	3	0.050	0.188	0.344	0.625	0.908	0.049	0.196	0.361	0.647	0.920

(0.2,0.7)	1	0.048	0.186	0.338	0.612	0.901	0.049	0.196	0.364	0.646	0.923
	2	0.050	0.190	0.344	0.617	0.904	0.050	0.199	0.366	0.647	0.924
	3	0.049	0.187	0.340	0.613	0.901	0.049	0.198	0.365	0.646	0.923
(0.1,0.9)	1	0.047	0.179	0.331	0.598	0.895	0.047	0.193	0.356	0.641	0.922
	2	0.053	0.189	0.340	0.606	0.897	0.050	0.198	0.361	0.643	0.922
	3	0.050	0.185	0.334	0.599	0.894	0.049	0.197	0.359	0.641	0.921
(0.5,0.5)	1	0.048	0.187	0.342	0.612	0.898	0.048	0.198	0.362	0.650	0.925
	2	0.049	0.190	0.346	0.619	0.904	0.048	0.199	0.364	0.652	0.926
	3	0.048	0.187	0.342	0.615	0.902	0.048	0.198	0.363	0.651	0.926
(0.9,0.9)	1	0.064	0.181	0.305	0.524	0.802	0.052	0.194	0.345	0.614	0.900
	2	0.059	0.182	0.313	0.544	0.827	0.052	0.194	0.347	0.618	0.903
	3	0.054	0.171	0.295	0.519	0.807	0.051	0.192	0.343	0.614	0.901

$n_1 = 19, n_2 = 29$							$n_1 = 19, n_2 = 19$				
		δ					δ				
$(\lambda_1, \ldots, \lambda_{10})$	Test[c]	0	2	4	8	16	0	2	4	8	16
$p = 10$											
(0.1,0.1,0.1,0.1,0.1, 0.1,0.1,0.1,0.1,0.1)	1	0.063	0.118	0.188	0.349	0.655	0.055	0.107	0.169	0.324	0.620
	2	0.056	0.111	0.179	0.343	0.657	0.058	0.113	0.178	0.342	0.646
	3	0.052	0.104	0.170	0.329	0.642	0.051	0.101	0.161	0.315	0.618
(0.1,0.1,0.1,0.1,0.5, 0.5,0.5,0.9,0.9,0.9)	1	0.049	0.099	0.163	0.328	0.652	0.046	0.094	0.153	0.308	0.617
	2	0.052	0.104	0.171	0.337	0.661	0.052	0.106	0.170	0.333	0.644
	3	0.052	0.104	0.171	0.338	0.661	0.047	0.098	0.158	0.315	0.626
(0.5,0.5,0.5,0.5,0.5, 0.5,0.5,0.5,0.5,0.5)	1	0.045	0.104	0.174	0.331	0.659	0.046	0.093	0.156	0.292	0.617
	2	0.043	0.100	0.170	0.331	0.660	0.052	0.104	0.171	0.316	0.645
	3	0.044	0.101	0.172	0.335	0.663	0.049	0.096	0.161	0.300	0.628
(0.1,0.1,0.1,0.1,0.1, 0.1,0.1,0.1,0.1,0.9)	1	0.053	0.112	0.189	0.339	0.661	0.074	0.129	0.192	0.338	0.612
	2	0.053	0.111	0.185	0.339	0.663	0.074	0.133	0.202	0.358	0.650
	3	0.051	0.108	0.182	0.333	0.658	0.057	0.106	0.165	0.306	0.591
(0.1,0.1,0.1,0.1,0.1, 0.5,0.5,0.5,0.5,0.5)	1	0.049	0.103	0.171	0.339	0.665	0.055	0.107	0.173	0.323	0.622
	2	0.048	0.102	0.170	0.338	0.665	0.058	0.114	0.183	0.340	0.647
	3	0.049	0.103	0.172	0.341	0.668	0.050	0.101	0.165	0.313	0.619
(0.1,0.2,0.3,0.4,0.5, 0.6,0.7,0.8,0.9,0.9)	1	0.051	0.104	0.175	0.339	0.662	0.047	0.093	0.153	0.309	0.619
	2	0.050	0.104	0.174	0.339	0.665	0.053	0.104	0.169	0.332	0.644
	3	0.051	0.104	0.174	0.339	0.665	0.048	0.096	0.158	0.316	0.626
(0.9,0.9,0.9,0.9,0.9, 0.9,0.9,0.9,0.9,0.9)	1	0.106	0.171	0.242	0.388	0.644	0.094	0.155	0.224	0.365	0.620
	2	0.083	0.145	0.217	0.376	0.664	0.083	0.144	0.217	0.372	0.655
	3	0.058	0.107	0.167	0.306	0.585	0.056	0.104	0.162	0.295	0.569

[a] 2=Johansen's test; 3=the MNV test.
[b] 1=Yao's test; 2=Johansen's test; 3=the MNV test.
[c] 1=Yao's test; 2=Johansen's test; 3=MNV test.

(5) For $p=10$, the only test that controls the sizes satisfactorily is the MNV test. In this case, the size of the Johansen's test goes up to 0.083 while the size of the MNV test never exceeds 0.058.

We also estimated the sizes and powers of these invariant tests for $p=3,4$ and 5. They are not reported here because these values exhibited similar pattern as those for the cases of $p=2$ and $p=10$. Power comparison between the MNV test and the one in Nel and Van der Merwe (1986) over the parameter configurations considered in Table 1 (the results are not reported here) showed that the MNV test is as powerful as Nel and Van der Merwe's test for smaller dimension, and slightly more powerful than the Nel and Van der Merwe test for larger dimension. On an overall basis, the MNV test seems to be the best among these three invariant tests. Our extensive numerical studies (results not reported here) showed that sizes of the MNV test are very close to the nominal level provided $n_1 \geqslant 5p$ and $n_2 \geqslant 5p$. This condition is somewhat relaxed for the equal sample size cases; in these cases, the required common sample size is at least $4p$.

4. Concluding remarks

In this article, we have shown numerically that the MNV solution is the best among available approximate invariant solutions for the multivariate Behrens–Fisher problem. The MNV method is satisfactory for practical purpose, and is simple to use. Furthermore, the MNV test can be easily extended to test or estimate any specified contrasts as in the one-sample case described in Seber (1984, Chapter 3). As we already pointed out, the available exact methods have serious drawbacks while satisfactory methods are approximate. As noted in Lehmann (1994, p. 304), a reasonable exact solution with natural properties may not exist. Therefore, there will be a continuous interest in developing better approximate solutions to the present problem. The MNV method may serve as a benchmark, and future tests may be compared with this test to establish their superiority.

Acknowledgements

The authors are grateful to a referee for providing valuable comments and suggestions.

Appendix

We here show that the degrees of freedom v in (7) satisfy that $\min\{n_1,n_2\} \leqslant v \leqslant n_1+n_2$. Let $l_1,\ldots,l_p$ denote the eigenvalues of $\tilde{S}_1\tilde{S}^{-1}$. Then it is easy to see that the degrees of freedom v in (7) can be written as

$$\frac{p+p^2}{(1/n_1)\left\{\sum_{i=1}^p l_i^2+\left(\sum_{i=1}^p l_i\right)^2\right\}+(1/n_2)\left\{\sum_{i=1}^p (1-l_i)^2+\left(\sum_{i=1}^p (1-l_i)\right)^2\right\}}. \tag{A.1}$$

Note that, for positive constants a and $b, (x^2/a) + ((1-x)^2/b)$ attains its minimum at $x = a/(a+b)$, and its minimum value is $1/(a+b)$. Using this fact, we see that

$$\frac{1}{n_1}\sum_{i=1}^{p} l_i^2 + \frac{1}{n_2}\sum_{i=1}^{p}(1-l_i)^2 \geqslant \frac{p}{n_1+n_2}.$$

Let $x = \sum_{i=1}^{p} l_i$. Then, it is easy to see that

$$\frac{1}{n_1}\left(\sum_{i=1}^{p} l_i\right)^2 + \frac{1}{n_2}\left(\sum_{i=1}^{p}(1-l_i)\right)^2 = \frac{x^2}{n_1} + \frac{(p-x)^2}{n_2} \geqslant \frac{p^2}{n_1+n_2}$$

and the equality holds when $x = pn_1/(n_1+n_2)$. The above two inequalities imply that the denominator of (A.1) is greater than or equal to $(p+p^2)/(n_1+n_2)$, and hence (A.1) is less than or equal to n_1+n_2.

To prove that $v \geqslant \min\{n_1, n_2\}$, let us assume without loss generality that $n_1 = \min\{n_1, n_2\}$. Using the fact that $0 \leqslant l_i \leqslant 1$ for $i = 1, \ldots, p$, it can be shown that the denominator of (A.1) is less than or equal to $(p+p^2)/n_1$. Therefore, (A.1) is greater than or equal to n_1.

References

Bennett, B.M., 1951. Note on a solution of the generalized Behrens–Fisher problem. Ann. Inst. Statist. Math. 2, 87–90.

Christensen, W.F., Rencher, A.C., 1997. A comparison of Type I error rates and power levels for seven solutions to the multivariate Behrens–Fisher problem. Comm. Statist.—Theory Methods 26, 1251–1273.

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.

Johansen, S., 1980. The Welch–James approximation to the distribution of the residual sum of squares in a weighted linear regression. Biometrika 67, 85–95.

Kim, S., 1992. A practical solution to the multivariate Behrens–Fisher problem. Biometrika 79, 171–176.

Lehmann, E.L., 1994. Testing Statistical Hypothesis. Chapman & Hall, New York.

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

Nel, D.G., Van der Merwe, C.A., 1986. A solution to the Multivariate Behrens–Fisher problem. Comm. Statist.—Theory Methods 15, 3719–3735.

Nel, D.G., Van der Merwe, C.A., Moser, B.K., 1990. The exact distributions of the univariate and multivariate Behrens–Fisher statistics with a comparison of several solutions in the univariate case. Comm. Statist.—Theory Methods 19, 279–298.

Seber, G.A.F., 1984. Multivariate Observations. Wiley, New York.

Smith, W.B., Hocking, R.R., 1972. Wishart variates generator (algorithm AS 53). Appl. Statist. 21, 341–345.

Subramaniam, K., Subramaniam, K., 1973. On the multivariate Behrens–Fisher problem. Biometrika 60, 107–111.

Tang, K.L., Algina, J., 1993. Performing of four multivariate tests under variance–covariance heteroscedasticity. Mult. Behavioral Res. 28, 391–405.

Welch, B.L., 1947. The generalization of Student's problem when several different population variances are involved. Biometrika 34, 28–35.

Yao, Y., 1965. An approximate degrees of freedom solution to the multivariate Behrens–Fisher problem. Biometrika 52, 139–147.