SOME ASPECTS OF MULTIVARIATE BEHRENS-FISHER PROBLEM

Junyong Park
Bimal Sinha
Department of Mathematics/Statistics
University of Maryland, Baltimore

Abstract

In this paper we discuss the well known multivariate Behrens-Fisher problem which deals with testing the equality of two normal mean vectors under heteroscedasticity of dispersion matrices. Some existing tests are reviewed and a new test based on Roy's union-intersection principle coupled with the generalized P-value is proposed. The tests are compared with respect to size and power based on simulation, and applied to a few useful data sets.

AMS Subject Classification: 62H10, 62H15
Keywords: Behrens-Fisher problem, generalized P-value, heteroscedasticity, MANOVA

1. Introduction

It is well known that the univariate Behrens-Fisher problem deals with statistical inference concerning the difference between the means of two univariate normal populations with unequal variances. Except for some simple *exact* but inefficient test procedures, most of the widely used solutions which have been recommended for the more general ANOVA problem under heteroscedasticty are approximate in nature (Hartung, Knapp and Sinha, 2006).

For the multivariate Behrens-Fisher problem which deals with testing the equality of two mean vectors under heteroscedastic dispersion matrices, finding reasonable solutions is even harder. Recently there has been some progress to solve this problem, generalizing some univariate approximate solutions and generalized P-value based exact solutions to the multivariate case.

In this paper we provide a complete review of all the existing methods and suggest a *new* test procedure, exploiting Roy's *union-intersection principle* (1957) coupled with Weerahandi's generalized P-value concept. We compare the size and power of some test procedures via simulation. A few applications are also included.

Section 2 contains notations and a survey of existing results. The new test is developed in Section 3. Simulation results are reported in Section 4 and applications are mentioned in Section 5.

2. Notations and existing results

Consider two p-variate normal populations $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ where μ_1 and μ_2 are unknown $p \times 1$ vectors and Σ_1 and Σ_2 are unknown $p \times p$ positive definite matrices. Let $\mathbf{X}_{\alpha 1} \sim N(\mu_1, \Sigma_1)$, $\alpha = 1, 2, ..., n_1$, and $\mathbf{X}_{\alpha 2} \sim N(\mu_2, \Sigma_2)$, $\alpha = 1, 2, ..., n_2$, denote random samples from these two populations, respectively. We are interested in the testing problem

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2. \tag{1}$$

For $i = 1, 2$, let

$$\bar{\mathbf{X}}_i = \frac{1}{n_i}\sum_{\alpha=1}^{n_i}\mathbf{X}_{\alpha i}, \tag{2}$$

$$\mathbf{A}_i = \sum_{\alpha=1}^{n_i}(\mathbf{X}_{\alpha i}-\bar{\mathbf{X}}_i)(\mathbf{X}_{\alpha i}-\bar{\mathbf{X}}_i)' \tag{3}$$

$$\mathbf{S}_i = \mathbf{A}_i/(n_i-1),\ \ i=1,\ 2. \tag{4}$$

Then $\bar{\mathbf{X}}_1$, $\bar{\mathbf{X}}_2$, $\mathbf{A}_1$ and $\mathbf{A}_2$, which are sufficient for the mean vectors and dispersion matrices, are independent random variables having the distributions:

$$\bar{\mathbf{X}}_i \sim N(\mu_i, \frac{\Sigma_i}{n_i}),\ \text{and } \mathbf{A}_i \sim W_p(n_i-1, \Sigma_i),\ \ i=1,\ 2 \tag{5}$$

where $W_p(r, \Sigma)$ denotes the $p-$dimensional Wishart distribution with df $= r$ and scale matrix Σ. Here is a brief review of most of the existing results in the literature for testing H_0 versus H_1.

Define

$$\tilde{\mathbf{S}}_i = \mathbf{S}_i/n_i,\ \ i=1,\ 2,\ \ \tilde{\mathbf{S}} = \tilde{\mathbf{S}}_1 + \tilde{\mathbf{S}}_2, T^2 = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)'\tilde{\mathbf{S}}^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2). \tag{6}$$

It should be noted that the test based on T^2 is a natural invariant test based on the union-intersection principle.

1. Yao (1965)'s invariant test. This is a multivariate extension of the Welch *approximate d.f.* solution, and is based on $T^2 \sim (\nu p/(\nu - p + 1))F_{p,\nu-p+1}$ with the $d.f.$ ν given by

$$\nu = \left[\frac{1}{n_1}\left(\frac{\bar{X}_d'\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}^{-1}\bar{X}_d}{\bar{X}_d\tilde{\mathbf{S}}^{-1}\bar{X}_d}\right) + \frac{1}{n_2}\left(\frac{\bar{X}_d'\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{S}}_2\tilde{\mathbf{S}}^{-1}\bar{X}_d}{\bar{X}_d\tilde{\mathbf{S}}^{-1}\bar{X}_d}\right)\right]^{-1} \tag{7}$$

2. Johansen (1980)'s invariant test (see also Tang and Algina, 1993). We use $T^2 \sim qF_{p,\nu}$ where

$$
\begin{aligned}
q &= p + 2D - 6D/[p(p-1)+2], \;\; \nu = p(p+2)/3D \qquad (8)\\
D &= \frac{1}{2}\sum_{i=1}^{2}\{tr[(I - (\tilde{\mathbf{S}}_1^{-1} + \tilde{\mathbf{S}}_2^{-1})^{-1}\tilde{\mathbf{S}}_i^{-1})^2] + tr[(I - (\tilde{\mathbf{S}}_1^{-1} + \tilde{\mathbf{S}}_2^{-1})^{-1}\tilde{\mathbf{S}}_i^{-1})]^2\}/n_i
\end{aligned}
$$

3. Nel and Van der Merwe (1986)'s *noninvariant* solution. Here we use $T^2 \sim (\nu p/(\nu - p + 1))F_{p,\nu-p+1}$ except that ν is defined by

$$
\nu_{NVM} = \frac{tr(\tilde{\mathbf{S}}^2) + [tr(\tilde{\mathbf{S}})]^2}{(1/n_1)\{tr(\tilde{\mathbf{S}}_1^2) + [tr(\tilde{\mathbf{S}}_1)]^2\} + (1/n_1)\{tr(\tilde{\mathbf{S}}_2^2) + [tr(\tilde{\mathbf{S}}_2)]^2\}} \qquad (9)
$$

4. Krishnamoorthy and Yu (2004)'s modified Nel/Van der Merwe invariant solution. We use the same idea as before, namely, $T^2 \sim (\nu p/(\nu - p + 1))F_{p,\nu-p+1}$ with the *d.f.* ν defined by $\nu_{KY} = (p + p^2)/C(\tilde{\mathbf{S}}_1, \tilde{\mathbf{S}}_2)$

$$
\begin{aligned}
C(\tilde{\mathbf{S}}_1, \tilde{\mathbf{S}}_2) &= \frac{1}{n_1}\{tr[(\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}^{-1})^2] + [tr(\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}^{-1})]^2\}\\
&\quad + \frac{1}{n_2}\{tr[(\tilde{\mathbf{S}}_2\tilde{\mathbf{S}}^{-1})^2] + [tr(\tilde{\mathbf{S}}_2\tilde{\mathbf{S}}^{-1})]^2\} \qquad (10)
\end{aligned}
$$

5. Kim (1992)'s test. Kim suggested a variation of T^2, given by

$$
\tilde{T}^2 = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)'[\tilde{\mathbf{S}}_1 + r^2\tilde{\mathbf{S}}_2 + 2rQ(\tilde{\mathbf{S}}_1, \tilde{\mathbf{S}}_2)]^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2). \qquad (11)
$$

where $r = [det(\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_2^{-1})]^{1/2p}$, and $Q(\tilde{\mathbf{S}}_1, \tilde{\mathbf{S}}_2) = \tilde{\mathbf{S}}_2^{1/2}[\tilde{\mathbf{S}}_2^{-1/2}\tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_2^{-1/2}]^{1/2}\tilde{\mathbf{S}}_2^{1/2}$. There are two drawbacks of Kim's (1992) test: it is *not* invariant and its type I error is below the nominal level most of the time.

6. Christensen and Rencher (1997). Based on extensive numerical results concerning the type I error and power, these authors recommend the approximate solution due to Kim (1992) mentioned above for testing (1) because many approximate tests have type I errors exceeding the nominal level.

7. Weerahandi (1995). The concepts of generalized $p-$values and generalized confidence regions were introduced by Tsui and Weerahandi (1989) and Weerahandi (1993), and applied to many nontrivial inference problems, including the Behrens-Fisher problem.

   Here is a brief review of the generalized p-value approach. Let X be a random variable whose distribution depends on the parameters (θ, δ), where θ is a scalar parameter of interest and δ represents nuisance parameters. Suppose we want to test $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, where θ_0 is a specified value.

   Let x denote the observed value of X and consider the *generalized test variable* $T(X; x, \theta, \delta)$, which also depends on the observed value and the parameters, and satisfies the following conditions:

   (a) The distribution of $T(X; x, \theta_0, \delta)$ is free of the nuisance parameter δ,

   (b) The observed value of $T(X; x, \theta_0, \delta)$, i.e., $T(x; x, \theta_0, \delta)$, is free of δ, and

   (c) $P[T(X; x, \theta, \delta) \geq t]$ is nondecreasing in θ, for fixed x and δ.

   Then the generalized $p-$value is defined by

   $$P[T(X; x, \theta_0, \delta) \geq t],$$

   where $t = T(x; x, \theta_0, \delta)$. It should be noted that, unlike the regular p values, the generalized p value does not follow a uniform distribution under the null hypothesis, and moreover, the type I error probability of a test based on the generalized $p-$value, and the coverage probability of a generalized confidence interval, may depend on the nuisance parameters. For further details and for applications, we refer to the book by Weerahandi (1995). For the univariate Behrens-Fisher problem, a test based on the generalized $p-$value is given in Tsui and Weerahandi (1989). For the multivariate Behrens-Fisher problem, an upper bound for the generalized $p-$value is given in Gamage (1997).

8. Gamage et al.(2004) explored the concept of the generalized $p-$value for testing the hypotheses in (1), and also for MANOVA problem involving

more than two normal populations. They were able to construct a generalized test variable whose observed value is $(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2})^{-1}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)$, a very natural quantity based on the sufficient statistics for the testing problem (1). Here $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ denote the observed values of $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$, respectively, and $\mathbf{s}_1$ and $\mathbf{s}_2$ denote the observed values of the sample covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ defined in (4). The computation of the generalized $p-$value as well as some numerical results on the type I error probabilities of the test based on the generalized $p-$value are reported in their paper. It turns out that the type I error probabilities are all below the nominal level. In other words, the test based on the generalized $p-$value, which is an exact probability of an extreme region, also provides a conservative test in the classical sense.

To describe the above generalized p-value based test procedure, define

$$
\begin{aligned}
\mathbf{Y}_1 &= \left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}\bar{\mathbf{X}}_1, \\
\mathbf{Y}_2 &= \left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}\bar{\mathbf{X}}_2, \\
\mathbf{V}_1 &= (\frac{n_1-1}{n_1})\left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}\mathbf{S}_1\left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}, \\
\mathbf{V}_2 &= (\frac{n_2-1}{n_2})\left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}\mathbf{S}_2\left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}, \\
\theta_1 &= \left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}\mu_1,\ \theta_2 = \left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}\mu_2, \\
\Lambda_1 &= \left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}\frac{\Sigma_1}{n_1}\left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}, \\
\Lambda_2 &= \left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}\frac{\Sigma_2}{n_2}\left(\frac{\mathbf{s}_1}{n_1}+\frac{\mathbf{s}_2}{n_2}\right)^{-1/2}
\end{aligned} \tag{12}
$$

where by $\mathbf{A}^{1/2}$ we mean the positive definite square root of the positive definite matrix $\mathbf{A}$, and $\mathbf{A}^{-1/2}=(\mathbf{A}^{1/2})^{-1}$. Then

$$\mathbf{Y}_1 \sim N(\theta_1,\Lambda_1),\ \mathbf{Y}_2\sim N(\theta_2,\Lambda_2), \tag{13}$$

$$\mathbf{V}_1 \sim W_p(n_1-1,\Lambda_1),\ \mathbf{V}_2\sim W_p(n_2-1,\Lambda_2) \tag{14}$$

and the testing problem (1) can be expressed as

$$H_0 : \theta_1 = \theta_2, \quad \text{vs} \quad H_1 : \theta_1 \neq \theta_2. \tag{15}$$

Let $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{v}_1$ and $\mathbf{v}_2$ denote the observed values of $\mathbf{Y}_1$, $\mathbf{Y}_2$, $\mathbf{V}_1$ and $\mathbf{V}_2$, respectively. Note that these observed values are obtained by replacing $\bar{\mathbf{X}}_1$, $\bar{\mathbf{X}}_2$, $\mathbf{S}_1$ and $\mathbf{S}_2$ by the corresponding observed values (namely, $\bar{\mathbf{x}}_1$, $\bar{\mathbf{x}}_2$, $\mathbf{s}_1$ and $\mathbf{s}_2$) in the expressions for $\mathbf{Y}_1$, $\mathbf{Y}_2$, $\mathbf{V}_1$ and $\mathbf{V}_2$ in (12). Then we have the distributions

$$\begin{aligned}
\mathbf{Z} &= (\Lambda_1 + \Lambda_2)^{-1/2}(\mathbf{Y}_1 - \mathbf{Y}_2) \sim N(0, I_p), \text{ under } H_0,\\
\mathbf{R}_1 &= \left[\mathbf{v}_1^{-1/2}\Lambda_1\mathbf{v}_1^{-1/2}\right]^{-1/2}\left[\mathbf{v}_1^{-1/2}\mathbf{V}_1\mathbf{v}_1^{-1/2}\right]\left[\mathbf{v}_1^{-1/2}\Lambda_1\mathbf{v}_1^{-1/2}\right]^{-1/2}\\
&\sim W_p(n_1 - 1, I_p),\\
\mathbf{R}_2 &= \left[\mathbf{v}_2^{-1/2}\Lambda_2\mathbf{v}_2^{-1/2}\right]^{-1/2}\left[\mathbf{v}_2^{-1/2}\mathbf{V}_2\mathbf{v}_2^{-1/2}\right]\left[\mathbf{v}_2^{-1/2}\Lambda_2\mathbf{v}_2^{-1/2}\right]^{-1/2}\\
&\sim W_p(n_2 - 1, I_p).
\end{aligned} \tag{16}$$

Now define

$$T_1 = \mathbf{Z}'\left[\mathbf{v}_1^{1/2}\mathbf{R}_1^{-1}\mathbf{v}_1^{1/2} + \mathbf{v}_2^{1/2}\mathbf{R}_2^{-1}\mathbf{v}_2^{1/2}\right]\mathbf{Z}. \tag{17}$$

Note that given $\mathbf{R}_1$ and $\mathbf{R}_2$, T_1 is a positive definite quadratic form in $\mathbf{Y}_1 - \mathbf{Y}_2$ (where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have the distributions in (13)), and T_1 is stochastically larger under H_1 than under H_0. Also, since the distributions of $\mathbf{Z}$, $\mathbf{R}_1$ and $\mathbf{R}_2$ are free of any unknown parameters (under H_0) and since these quantities are independent, it follows that the distribution of T_1 is free of any unknown parameters (under H_0). Using the definition of $\mathbf{Z}$, $\mathbf{R}_1$ and $\mathbf{R}_2$, we conclude that

$$\begin{aligned}
\text{the observed value of } T_1 &= (\mathbf{y}_1 - \mathbf{y}_2)'(\mathbf{y}_1 - \mathbf{y}_2)\\
&= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\left(\frac{\mathbf{s}_1}{n_1} + \frac{\mathbf{s}_2}{n_2}\right)^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\\
&= t_1 \text{ (say)},
\end{aligned} \tag{18}$$

which also does not depend on any unknown parameters. In other words, T_1 is a generalized test variable. Hence the generalized p-value is given by

$$P(T_1 \geq t_1 | H_0) \tag{19}$$

where T_1 and t_1 are given in (17) and (18), respectively.

Computation of the generalized P-value based on a suitable representation of T_1 is discussed in Gamage et al. (2004).

9. Gamage et al. (2004) suggested another test for the multivariate Behrens-Fisher problem, which is different from the one mentioned above. This is based on the matrix identity:

$$(n_1\Sigma_1^{-1} + n_2\Sigma_2^{-1})^{-1} = \frac{\Sigma_1}{n_1}\left(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}\right)^{-1}\frac{\Sigma_2}{n_2} = \frac{\Sigma_2}{n_2}\left(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}\right)^{-1}\frac{\Sigma_1}{n_1},$$

resulting in the generalized test variable T_2 given by

$$T_2 = \frac{(\bar{X}_1 - \bar{X}_2)'(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2})^{-1}(\bar{X}_1 - \bar{X}_2)}{(\bar{x}_1 - \bar{x}_2)'\left[\frac{1}{n_1}s_1^{1/2}R_1^{*-1}s_1^{1/2} + \frac{1}{n_2}s_2^{1/2}R_2^{*-1}s_2^{1/2}\right]^{-1}(\bar{x}_1 - \bar{x}_2)}. \tag{20}$$

The generalized p-value based on the generalized test variable T_2 in (20) is given by

$$P\left(T_2 \geq 1 \middle| H_0\right). \tag{21}$$

An unfortunate feature of this generalized p-value is that it is not invariant under nonsingular transformations, unlike the generalized p-value defined earlier in (19).

3. A new test procedure

To describe the new test based an an application of Roy's union-intersection principle and Weerahandi's generalized P-value idea, let us first look at the familiar Hotelling's T^2 test dealing with a test for the mean vector of a single multivariate normal population $N[\mu, \Sigma]$. Denoting by $\bar{\mathbf{X}}$ the sample mean vector and by $\mathbf{S}$ the sample Wishart matrix, Hotelling's T^2 is essentially given by

$$T^2 = (\bar{\mathbf{X}} - \mu)\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu) \tag{22}$$

and the test for μ rejects the null hypothesis when T^2 exceeds its observed value

$$T^2_{obs} = (\bar{\mathbf{x}} - \mu)'\mathbf{s}^{-1}(\bar{\mathbf{x}} - \mu) \tag{23}$$

where $\bar{\mathbf{x}}$ and $\mathbf{s}$ are the observed entities.

Roy's derivation of T^2 is based on the univariate t test for a linear function of μ, say $\mathbf{a}'\mu$, using

$$t^2(\mathbf{a}) = [\mathbf{a}'(\bar{\mathbf{X}} - \mu)]^2/[\mathbf{a}'\mathbf{S}\mathbf{a}] \tag{24}$$

and rejecting a hypothesis for μ when $t^2(\mathbf{a})$ exceeds its observed value

$$t^2_{obs}(\mathbf{a}) = [\mathbf{a}'(\bar{\mathbf{x}} - \mu)]^2/[\mathbf{a}'\mathbf{s}\mathbf{a}]. \tag{25}$$

Hotelling's T^2 test based on (22) and (23) then follows upon combining such tests from (24) and (25) for *all* nonnull vectors $\mathbf{a}$, resulting in the P-value

$$P = P[sup_{\mathbf{a}\neq\mathbf{0}} t^2(\mathbf{a}) > sup_{\mathbf{a}\neq\mathbf{0}} t^2_{obs}(\mathbf{a})]. \tag{26}$$

We are now in a position to describe the new test for the multivariate Behrens-Fisher problem. Towards this end, we consider Weerahandi's generalized P-value solution for the univariate Behrens-Fisher problem and then apply Roy's principle as described above to extend the solution from the univariate case to the multivariate case.

Consider two univariate normal populations $N[\mu_1, \sigma_1^2]$ and $N[\mu_2, \sigma_2^2]$ and the problem of drawing inference about $\mu_1 - \mu_2$ based on random samples of sizes n_1 and n_2 from the two populations, respectively. Denoting the sample means and sample sums of squares by $(\bar{X}_1, \bar{X}_2, S_1^2, S_2^2)$, and their observed values by $(\bar{x}_1, \bar{x}_2, s_1^2, s_2^2)$, Weerahandi's (1995) test variable $T(.)$ is given by

$$\begin{aligned} & T(\bar{X}_1, \bar{X}_2, S_1^2, S_2^2; \bar{x}_1, \bar{x}_2, s_1^2, s_2^2; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \\ = & (\bar{X}_1 - \bar{X}_2) \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)^{-1/2} \left(\frac{\sigma_1^2 \, s_1^2}{S_1^2 n_1} + \frac{\sigma_2^2 s_2^2}{S_2^2 n_2} \right)^{1/2} \end{aligned} \tag{27}$$

with its observed value as $T_{obs} = \bar{x}_1 - \bar{x}_2$. The generalized P-value against one-sided alternative is then obtained from

$$\begin{aligned} P &= P(T \geq \bar{x}_1 - \bar{x}_2) \\ &= P\left[Z \left(\frac{1}{U_1} \frac{s_1^2}{n_1} + \frac{1}{U_2} \frac{s_2^2}{n_2} \right)^{1/2} \geq \bar{x}_1 - \bar{x}_2 \right] \end{aligned} \tag{28}$$

where $Z \sim N(0,1)$, $U_1 = S_1^2/\sigma_1^2 \sim \chi^2_{n_1-1}$ and $U_2 = S_2^2/\sigma_2^2 \sim \chi^2_{n_2-1}$.

For both-sided alternatives, the generalized P value is given by

$$P = P\left[Z^2 \left(\frac{1}{U_1} \frac{s_1^2}{n_1} + \frac{1}{U_2} \frac{s_2^2}{n_2} \right) \geq (\bar{x}_1 - \bar{x}_2)^2 \right] \tag{29}$$

which can also be expressed as

$$P = P\left[\frac{(\bar{X}_1 - \bar{X}_2)^2}{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})} \geq \frac{(\bar{x}_1 - \bar{x}_2)^2}{(\frac{s_1^2}{U_1 n_1} + \frac{s_2^2}{U_2 n_2})} \right]. \tag{30}$$

Returning to the multivariate case, we select a linear function $\mathbf{a}'(\mu_1 - \mu_2)$ and apply the above test variable $T(.)$ defined in (27) based on $(\mathbf{a}'\bar{\mathbf{X}}_1, \mathbf{a}'\bar{\mathbf{X}}_2, \mathbf{a}'\mathbf{S}_1\mathbf{a}, \mathbf{a}'\mathbf{S}_2\mathbf{a})$ and its observed value $(\mathbf{a}'\bar{\mathbf{x}}_1, \mathbf{a}'\bar{\mathbf{x}}_2, \mathbf{a}'\mathbf{s}_1\mathbf{a}, \mathbf{a}'\mathbf{s}_2\mathbf{a})$, resulting in $T(\mathbf{a})$ and hence the P-value, $P(\mathbf{a})$ as

$$\begin{aligned} P(\mathbf{a}) \;=\; P\Big[& (\mathbf{a}'\bar{\mathbf{X}}_1 - \mathbf{a}'\bar{\mathbf{X}}_2)^2 (\frac{\mathbf{a}'\Sigma_1\mathbf{a}}{n_1} + \frac{\mathbf{a}'\Sigma_2\mathbf{a}}{n_2})^{-1} \\ & \geq (\mathbf{a}'\bar{\mathbf{x}}_1 - \mathbf{a}'\bar{\mathbf{x}}_2)^2 (\frac{\mathbf{a}'\Sigma_1\mathbf{a}\mathbf{a}'\mathbf{s}_1\mathbf{a}}{\mathbf{a}'\mathbf{S}_1\mathbf{a}n_1} + \frac{\mathbf{a}'\Sigma_2\mathbf{a}\mathbf{a}'\mathbf{s}_2\mathbf{a}}{\mathbf{a}'\mathbf{S}_2\mathbf{a}n_2})^{-1} \Big]. \end{aligned} \tag{31}$$

We now use the fact that for any $\mathbf{a} \neq \mathbf{0}$, $\mathbf{a}'\mathbf{S}_i\mathbf{a}/\mathbf{a}'\Sigma_i\mathbf{a} = U_i \sim \chi^2_{n_i-1}$, $i = 1, 2$, and

$$sup_{\mathbf{a}\neq\mathbf{0}}(\mathbf{a}'\bar{\mathbf{X}}_1 - \mathbf{a}'\bar{\mathbf{X}}_2)^2(\frac{\mathbf{a}'\Sigma_1\mathbf{a}}{n_1} + \frac{\mathbf{a}'\Sigma_2\mathbf{a}}{n_2})^{-1} \quad (32)$$
$$= (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)'(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2})^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$$
$$\sim \chi^2_p$$

and

$$sup_{\mathbf{a}\neq\mathbf{0}}\frac{(\mathbf{a}'\bar{\mathbf{x}}_1 - \mathbf{a}'\bar{\mathbf{x}}_2)^2}{(\frac{\mathbf{a}'s_1\mathbf{a}}{U_1n_1} + \frac{\mathbf{a}'s_2\mathbf{a}}{U_2n_2})} \quad (33)$$
$$= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'(\frac{s_1}{U_1n_1} + \frac{s_2}{U_2n_2})^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2).$$

Finally, applying Roy's principle, we get the generalized P-value of the new test procedure as

$$P\left[\chi^2_p > (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'(\frac{s_1}{U_1n_1} + \frac{s_2}{U_2n_2})^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\right] \quad (34)$$

where $U_1 \sim \chi^2_{n_1-1}$ and $U_2 \sim \chi^2_{n_2-1}$. It is interesting to note the similarity of the new test procedure with the one suggested in Gamage et al. (2004), although carrying out our test is much easier. Simulation results showing the size and power of this and other tests are given in the next section.

4. Simulation

Simlations studies are based on Krishnamoorthy and Yu (2004). We compute the size and the power when $p = 2$. $\tilde{\Sigma}_1 = \Lambda$ and $\tilde{\Sigma}_2 = I - \Lambda$ where $\tilde{\Sigma}_i = \Sigma_i/n_i$ and $\Lambda = diag(\lambda_1, \lambda_2)$. The noncentrality parameter is $\delta = (\mu_1 - \mu_2)'(\tilde{\Sigma}_1 + \tilde{\Sigma}_2)(\mu_1 - \mu_2)$ and the mean vectors are chosen such that $\mu_1 - \mu_2 = (\sqrt{\delta/p})\mathbf{1}$ where $\mathbf{1}$ is the vector of ones. δ is taken as $0, 2, 4, 8, 16$, and $(n_1, n_2) = (6, 12)$ and $(12, 12)$ as in Krishnamoorthy and Yu (2004). We consider 6 tests in Table 1 and Table 2, namely Tests 1-6; Test 1 : Yao's test(1965), Test 2 : Johansen's test(1980), Test 3 : Krishnamoorthy and Yu's test(2004), Test 4 : T_1 in Gamage et al.(2004), Test 5 : T_2 in Gamage et al.(2004) and Test 6 : Proposed new test.

Table 1: $n_1 = 6$ and $n_2 = 12$

		δ				
(λ_1, λ_2)	Test	0	2	4	8	16
(0.1,0.1)	Test1	0.051	0.191	0.341	0.612	0.900
	Test2	0.051	0.193	0.345	0.617	0.903
	Test3	0.050	0.190	0.340	0.612	0.901
	Test4	0.033	0.157	0.286	0.545	0.848
	Test5*	0.130	0.346	0.534	0.798	0.966
	Test6	0.040	0.146	0.289	0.563	0.871
(0.2,0.5)	Test1	0.049	0.186	0.342	0.623	0.907
	Test2	0.051	0.190	0.347	0.628	0.910
	Test3	0.050	0.188	0.344	0.625	0.908
	Test4	0.022	0.116	0.216	0.475	0.817
	Test5*	0.095	0.286	0.480	0.750	0.956
	Test6	0.024	0.129	0.240	0.497	0.832
(0.2,0.7)	Test1	0.048	0.186	0.338	0.612	0.901
	Test2	0.050	0.190	0.344	0.617	0.904
	Test3	0.049	0.187	0.340	0.613	0.901
	Test4	0.028	0.112	0.211	0.442	0.795
	Test5*	0.096	0.300	0.484	0.738	0.952
	Test6	0.030	0.133	0.239	0.504	0.824
(0.1,0.9)	Test1	0.047	0.179	0.331	0.598	0.895
	Test2	0.053	0.189	0.340	0.606	0.897
	Test3	0.050	0.185	0.334	0.599	0.894
	Test4	0.028	0.108	0.213	0.400	0.770
	Test5*	0.096	0.308	0.476	0.744	0.954
	Test6	0.037	0.153	0.266	0.524	0.852
(0.5,0.5)	Test1	0.048	0.187	0.342	0.612	0.898
	Test2	0.049	0.190	0.346	0.619	0.904
	Test3	0 048	0.187	0.342	0.615	0.902
	Test4	0.018	0.111	0.216	0.430	0.778
	Test5*	0.102	0.296	0.474	0.732	0.948
	Test6	0.020	0.125	0.233	0.474	0.804
(0.9,0.9)	Test1	0.064	0.181	0.350	0.524	0.802
	Test2	0.059	0.182	0.313	0.544	0.827
	Test3	0.054	0.171	0.295	0.519	0.807
	Test4	0.037	0.124	0.224	0.423	0.675
	Test5*	0.148	0.364	0.506	0.738	0.952
	Test6	0.059	0.169	0.286	0.468	0.759

Table 2: $n_1 = n_2 = 12$

(λ_1, λ_2)	Test	δ 0	2	4	8	16
(0.1,0.1)	Test1	0.053	0.193	0.347	0.617	0.900
	Test2	0.052	0.194	0.350	0.621	0.904
	Test3	0.051	0.191	0.346	0.617	0.901
	Test4	0.052	0.190	0.368	0.635	0.906
	Test5*	0.141	0.358	0.562	0.809	0.967
	Test6	0.049	0.189	0.356	0.625	0.905
(0.2,0.5)	Test1	0.048	0.196	0.360	0.646	0.920
	Test2	0.049	0.197	0.362	0.648	0.921
	Test3	0.049	0.196	0.361	0.647	0.920
	Test4	0.040	0.192	0.373	0.668	0.909
	Test5*	0.120	0.324	0.535	0.794	0.972
	Test6	0.032	0.169	0.333	0.634	0.903
(0.2,0.7)	Test1	0.049	0.196	0.364	0.646	0.923
	Test2	0.050	0.199	0.366	0.647	0.924
	Test3	0.049	0.198	0.365	0.646	0.923
	Test4	0.039	0.186	0.360	0.656	0.911
	Test5*	0.117	0.334	0.522	0.791	0.972
	Test6	0.033	0.169	0.330	0.639	0.905
(0.1,0.9)	Test1	0.047	0.193	0.356	0.641	0.922
	Test2	0.050	0.198	0.361	0.643	0.922
	Test3	0.049	0.197	0.359	0.641	0.921
	Test4	0.038	0.169	0.332	0.625	0.897
	Test5*	0.117	0.349	0.515	0.791	0.975
	Test6	0.041	0.179	0.348	0.644	0.910
(0.5,0.5)	Test1	0.048	0.198	0.362	0.650	0.925
	Test2	0.048	0.199	0.364	0.652	0.926
	Test3	0.048	0.198	0.363	0.651	0.926
	Test4	0.036	0.209	0.378	0.671	0.915
	Test5*	0.108	0.332	0.524	0.788	0.966
	Test6	0.027	0.171	0.315	0.612	0.902
(0.9,0.9)	Test1	0.052	0.194	0.345	0.614	0.900
	Test2	0.052	0.194	0.347	0.618	0.903
	Test3	0.051	0.192	0.343	0.614	0.901
	Test4	0.051	0.217	0.360	0.729	0.901
	Test5*	0.133	0.373	0.558	0.815	0.976
	Test6	0.048	0.206	0.348	0.621	0.898

We conclude from the two tables that most of the tests except the non-invariant Test 5 proposed by Gamage et al. (2004) are quite reasonable in terms of both size and power.

5. Two applications

In this section, we consider two applications of the preceding tests. Our first application involves comparing two types of coating for resistance to corrosion based on paired data of size 15. The data is taken from Kramer and Jensen (1969) and also reported in Rencher (2002). Here is the data.

Table 3: Depth of Maximum Pits

	Coating 1		Coating 2	
	Depth	Number	Depth	Number
Location	y_1	y_2	x_1	x_2
1	73	31	51	35
2	43	19	41	14
3	47	22	43	19
4	53	26	41	29
5	58	36	47	34
6	47	30	32	26
7	52	29	24	19
8	38	36	43	37
9	61	34	53	24
10	56	33	52	27
11	56	19	57	14
12	34	19	44	19
13	55	26	57	30
14	65	15	40	7
15	75	18	68	13

The 6 tests applied on this data set result in the following p-values.

Again, except Test 5, all the other tests result in the acceptance of the null hypothesis.

Table 4: p-values for corrosion data

Test	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
p-value	0.1072	0.1047	0.1076	0.1065	0.0428	0.1358

Our second example deals with student's essay data taken from Kramer(1972) and also reported in Rencher (2002). The variables recorded were the number of words and the number of verbs for 15 students, and the testing problem was to compare the mean performances under formal and informal instructions.

Table 5: Number of Words and Number of Verbs

	Informal		Formal	
	Words	Verbs	Words	Verbs
Student	y_1	y_2	x_1	x_2
1	148	20	137	15
2	159	24	164	25
3	144	19	224	27
4	103	18	208	33
5	121	17	178	24
6	89	11	128	20
7	119	17	154	18
8	123	13	158	16
9	76	16	102	21
10	217	29	214	25
11	148	22	209	24
12	151	21	151	16
13	83	7	123	13
14	135	20	161	22
15	178	15	175	23

Based on the p-values reported in Table 6, we conclude that all tests except Test 5 accept the null hypothesis.

Table 6: p-values for essay data

Test	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
p-value	0.0960	0.0720	0.0721	0.0670	0.0293	0.0954

References

1. W. F. Christensen and A. C. Rencher, A comparison of type I error rates and power levels for seven solutions to the multivariate Beherns-Fisher problem, *Communications in Statist. − Simulation and Computation*, **26** (1997), 1251-1273.

2. J. Gamage, Generalized p-values and the multivariate Beherns-Fisher Problem, *Linear Algebra and Its Applications*, **253** (1997), 369-377.

3. J. Gamage and S. Weerahandi, Size performance of some tests in one-way ANOVA, *Communications in Statist. − Simulation and Computation*, **27** (1998), 625-640.

4. J. Gamage, T. Mathew, and S. Weerahandi, Generalized p-values and generalized condifence regions for the multivariate Behrens-Fisher problem and MANOVA *Journal of Multivariate Analysis*, **88** (2004), 177-189.

5. J. Hartung, G. Knapp and B. Sinha, Statistical Methods of Meta Analysis (2006).

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

7. S. Kim, A practical solution to the multivariate Beherns-Fisher problem, *Biometrika*, **79** (1992), 171-176.

8. C.Y. Kramer, A First course in Multivariate Analysis, Published by the author, Blacksburg, VA (1972)

9. C.Y. Kramer and D.R. Jensen, Fundamentals of Multivariate Analysis, Part II. Inference about Two Treatments, *Journal of Quality of Technology*, **1** (1969), 189-204.

10. K. Krishnamoorthy and J. Yu, Modified Nel and Van der Merwe test for the multivariate Behrens-Fisher problem, *Statistica and Probability Letters*, **66** (2004), 161-169.

11. D.G. Nel and C.A. Van der Merwe, A solution to the multivariate Behrens-Fisher problem, *Communications in Statistics:Theory and Methods* **15** (1986), 3719-3735.

12. A.C. Rencher, Methods of Multivariate Analysis second edition, John Wiley (2002).

13. S.N. Roy, Some Aspects of Multivariate Analysis, John Wiley and Sons, New York (1957).

14. K. W. Tsui and S. Weerahandi, Generalized p-values in siginificance testing of hypotheses in the presence of nuisance parameters, *J. Amer. Statist. Assoc.*, **84** (1989), 602-607.

15. S. Weerahandi, Generalized confidence intervals, *J. Amer. Statist. Assoc.*, **88** (1993), 899-905.

16. S. Weerahandi, "Exact Statistical Methods for Data Analysis," Springer-Verlag, New York, 1995.

17. Y. Yao, An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem, *Biometrika* (1965), **52**, 139-147.