Statistica Sinica **6**(1996), 419-431

AN IDENTITY FOR THE NONCENTRAL MULTIVARIATE F DISTRIBUTION WITH APPLICATION

Pui Lam Leung and Milton Lo

The Chinese University of Hong Kong

Abstract. Muirhead and Verathaworn (1985) and Konno (1991a,b) extended the Wishart identity to the multivariate F distribution and used this identity to prove dominance results for estimating the scale matrix of the multivariate F distribution. This paper extends this F identity to the noncentral multivariate F distribution. As an application of this noncentral F identity, we consider the problem of estimating the noncentrality matrix of a noncentral multivariate F distribution. This identity is used to develop a class of orthogonally invariant estimators which dominate the usual unbiased estimator. A simulation study was carried out to compare the performance of these estimators.

Key words and phrases: Wishart identity, noncentrality matrix, decision–theoretic estimation, noncentral multivariate F distribution.

1. Introduction

In estimating the latent roots in a two sample setting, Muirhead and Verathaworn (1985) derived an identity for the multivariate F distribution. This F identity is similar to the Wishart identity (derived independently by C. Stein and L. Haff) and is very useful in finding the risk difference between estimators. Later, Konno (1991a, 1992) and Leung (1992) used this F identity in estimating the scale matrix of the multivariate F distribution. For reference, we define the following symbols and state this F identity.

Suppose that a random $m \times m$ positive definite matrix $F = (f_{ij})$ has a multivariate F distribution with degrees of freedom n_1 and n_2 and scale matrix Ω, denoted by $F_m(n_1, n_2; \Omega)$. That is, F has the probability density function

$$\frac{\Gamma_m[(n_1+n_2)/2]}{\Gamma_m(n_1/2)\Gamma_m(n_2/2)}\ (\det\Omega)^{-n_1/2}(\det F)^{(n_1-m-1)/2}[\det(I+\Omega^{-1}F)]^{-n/2},$$

where $n = n_1 + n_2$, $n_1 > m+1$, $n_2 > m+1$ and $\Gamma_m(\cdot)$ is the multivariate Gamma function. Let $V(F,\Omega)$ be a matrix whose elements are function of F and Ω and let $V_{(r)} = rV + (1-r)\operatorname{diag}(V)$. Define

$$D = (d_{ij}) = \frac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial f_{ij}} \tag{1.1}$$

as a matrix of differential operators where δ_{ij} is the Kronecker delta. DV is the formal matrix product of D and V. Let $h(F)$ be a real–valued function of F and $\partial h(F)/\partial F = (\partial h(F)/\partial f_{ij})$. Write $V(F,\Omega)$ as V and $h(F)$ as h for brevity. Under fairly general regularity conditions, we have the F identity :

$$E[h\,\mathrm{tr}(\Omega+F)^{-1}V] = \frac{2}{n}E[h\,\mathrm{tr}(DV)] + \frac{2}{n}E\Big[\mathrm{tr}(\frac{\partial h}{\partial F}V_{(1/2)})\Big] + \frac{n_1-m-1}{n}E[h\,\mathrm{tr}(F^{-1}V)]. \quad (1.2)$$

This F identity is an extension of the Wishart identity to the multivariate F distribution. The regularity conditions (to ensure the function hV satisfies the conditions of the Stokes' theorem) are given in Konno (1988).

The Wishart identity was used to prove dominance results in decision-theoretic estimation problems in a series of papers Haff (1979a,b, 1980, 1981, 1982). It is natural to look for a similar identity for the corresponding non-central distributions. Leung (1994a) generalized the Wishart identity to the noncentral Wishart distribution and applied this noncentral Wishart identity to an estimation problem. In the present paper, we generalize the F identity (1.2) to the noncentral F distribution, and the result will be called the "noncentral F identity". In Section 2, the noncentral F identity is stated and proved. As an application of this identity, we consider the problem of estimating the noncentrality matrix Δ of a noncentral multivariate F distribution in Section 3. A class of orthogonally invariant estimator of Δ is proposed which dominates the usual unbiased estimator of Δ. A simulation study was carried out to compare the performance of the proposed estimator. This problem has also been considered by Leung and Muirhead (1987).

2. The Noncentral F Identity

Suppose that $m\times m$ matrices A and B are independent with noncentral Wishart and central Wishart distributions respectively. A has n_1 degrees of freedom, identity covariance matrix and noncentrality matrix Δ, denoted by $W_m(n_1,I,\Delta)$. B has n_2 degrees of freedom and identity covariance matrix, denoted by $W_m(n_2,I)$. Define $F = A^{1/2}B^{-1}A^{1/2}$; then F has a noncentral multivariate F distribution, denoted by $F_m(n_1,n_2;I;\Delta)$, with probability density function

$$g(F) = \frac{\Gamma_m(n/2)\mathrm{etr}(-\Delta/2)}{\Gamma_m(n_1/2)\Gamma_m(n_2/2)}\frac{(\det F)^{(n_1-m-1)/2}}{[\det(I+F)]^{n/2}}\,{}_1F_1[\frac{n}{2};\frac{n_1}{2};\frac{1}{2}\Delta F(I+F)^{-1}],$$

where $\mathrm{etr}(\cdot) = \exp[\mathrm{tr}(\cdot)]$, ${}_1F_1(\cdot)$ is the confluent Hypergeometric function with matrix argument and $n = n_1+n_2$ (see Muirhead (1982) for details). Under the same regularity conditions for the F identity given in Konno (1988), we have

Theorem 2.1. (Noncentral F identity)

$$E[h\,\mathrm{tr}(I+F)^{-1}V]=\frac{2}{n}E[h\,\mathrm{tr}(DV)]+\frac{2}{n}E\Big[\mathrm{tr}(\frac{\partial h}{\partial F}V_{(1/2)})\Big] \qquad (2.1)$$
$$+\frac{n_1-m-1}{n}E[h\,\mathrm{tr}(F^{-1}V)]+\frac{1}{n}E_1\{h\,\mathrm{tr}[F^{-1}\Delta(I+F)^{-1}V]\},$$

where the expectation E is taken over a $F_m(n_1,n_2;I;\Delta)$ distribution and

$$E_1\{h\,\mathrm{tr}[F^{-1}(I+F)^{-1}V]\}=\int_{F>0}h\,\mathrm{tr}[F^{-1}(I+F)^{-1}V]g_1(F)(dF)$$

with $g_1(F)$ the density of a $F_m(n_1+m+1,n_2;I;\Delta)$ distribution.

The identity (2.1) is same as (1.2) except for the last term where the expectation is taken over a noncentral multivariate F distribution with degrees of freedom changed from n_1 to n_1+m+1. Notice that when $\Delta=0$, (2.1) reduces to (1.2) with $\Omega=I$. Before proving (2.1), we need the following lemmas.

Lemma 2.2.
(i) $E[F]=(n_1I+\Delta)/c_1$,
(ii) $E[\mathrm{tr}(F^2)]=[(\mathrm{tr}\,\Delta)^2+c_1\,\mathrm{tr}(\Delta^2)+c_2\,\mathrm{tr}\,\Delta+c_3]/c_0$,
(iii) $E[(\mathrm{tr}\,F)^2]=[c_4(\mathrm{tr}\,\Delta)^2+2\,\mathrm{tr}(\Delta^2)+c_5(\mathrm{tr}\,\Delta)+c_6]/c_0$,
where $c_0=(n_2-m)(n_2-m-1)(n_2-m-3)$, $c_1=n_2-m-1$, $c_2=2[(n_2-m)(n_1+m+1)+(m-1)(n_1-1)]$, $c_3=mn_1c_2/2$, $c_4=n_2-m-2$, $c_5=2[c_4(mn_1+2)+2(n_1+m+1)]$, $c_6=mn_1c_5/2$.

Proof. (i) can be easily proved by the conditioning on A as follows:

$$E[F]=E[A^{1/2}B^{-1}A^{1/2}]=E[E(A^{1/2}B^{-1}A^{1/2}|A)]$$
$$=\frac{E(A)}{n_2-m-1}=\frac{n_1I+\Delta}{n_2-m-1}.$$

The proofs of (ii) and (iii) are reasonably straightforward (but messy) using the conditioning on A as in (i) and Theorem 4.4 of Magnus and Neudecker (1979) (see Leung (1986) for details). Note that the result (ii) is also given in Lemma 3.3 of Leung and Muirhead (1987).

We also need formulae for the differential operator D (defined in (1.1)) on F^2 and F^3.

Lemma 2.3.
(i) $\mathrm{tr}[DF^2]=(m+1)\,\mathrm{tr}\,F$.
(ii) $\mathrm{tr}[DF^3]=\frac{2m+3}{2}\,\mathrm{tr}(F^2)+\frac{1}{2}(\mathrm{tr}\,F)^2$.

Proof. (i) and (ii) can be easily obtained using Lemma 2.3 in Konno (1991b) and is omitted.

Now we turn to the proof of Theorem 2.1.

Proof of (2.1). The proof is very similar to the proof of Theorem 2.1 in Leung (1994a). From above, the density of F is

$$g(F) = C\frac{(\det F)^{(n_1-m-1)/2}}{[\det(I+F)]^{n/2}}\,{}_1F_1\Big[\frac{n}{2};\frac{n_1}{2};\frac{1}{2}\Delta F(I+F)^{-1})\Big], \qquad (2.2)$$

where $C=[\Gamma_m(n/2)\mathrm{etr}(-\Delta/2)]/[\Gamma_m(n_1/2)\Gamma_m(n_2/2)]$. For the differential operator D defined in (1.1), we have $D(\det F)^\alpha = \alpha(\det F)^\alpha F^{-1}$. Hence, D operating on $g(F)$ in (2.2) gives

$$Dg(F) = \Big[\frac{n_1-m-1}{2}F^{-1} - \frac{n}{2}(I+F)^{-1}\Big]g(F) + Q(F), \qquad (2.3)$$

where

$$Q(F) = C\frac{(\det f)^{(n_1-m-1)/2}}{[\det(I+F)]^{n/2}}\Big\{D_1F_1[\frac{n}{2};\frac{n_1}{2};\frac{1}{2}\Delta F(I+F)^{-1}]\Big\}. \qquad (2.4)$$

The same set of regularity conditions on hV given in Konno (1988) ensures that

$$\int_{F>0}\mathrm{tr}\, D[hVg(F)](dF) = 0.$$

It follows that

$$0 = E\,\mathrm{tr}[(\partial h/\partial F)V_{(1/2)}] + E[h\,\mathrm{tr}(DV)] + \int_{F>0} hV\,\mathrm{tr}[Dg(F)](dF).$$

Using (2.3), we have

$$\begin{aligned} E[h\,\mathrm{tr}(I+F)^{-1}V] = {} & \frac{2}{n}E[h\,\mathrm{tr}(DV)] + \frac{2}{n}E\Big[\mathrm{tr}(\frac{\partial h}{\partial F}V_{(1/2)})\Big] \\ & + \frac{n_1-m-1}{n}E[h\,\mathrm{tr}(F^{-1}V)] + \frac{2}{n}\int_{F>0} h\,\mathrm{tr}(QV)(dF). \end{aligned} \qquad (2.5)$$

Comparing (2.5) with (2.1), the proof is complete if we can show

$$2\int_{F>0} h\,\mathrm{tr}(QV)(dF) = \int_{F>0} h\,\mathrm{tr}[F^{-1}\Delta(I+F)^{-1}V]g_1(F)(dF),$$

where $g_1(F)$ is the density of a $F_m(n_1+m+1, n_2; I; \Delta)$ distribution. Therefore, it suffices to show that

$$2Q\ g_1^{-1}(F) = F^{-1}\Delta(I+F)^{-1} \qquad \text{a.e.} \qquad (2.6)$$

$Q(F)$ defined in (2.4) involves the operation of D on ${}_1F_1(\cdot)$. Although it is possible to prove (2.6) directly by differentiating the series of zonal polynomials in ${}_1F_1(\cdot)$, this could be very complicated and messy. We take another approach. By using $V = (I+F)F^2$ and $h = 1$ in (2.5) with Lemma 2.3 and simplifying, we obtain

$$E[\mathrm{tr}(F^2)] = \frac{n_1+m+1}{n_2-m-2}E[\mathrm{tr}\,F] + \frac{1}{n_2-m-2}E[(\mathrm{tr}\,F)^2] + \frac{2}{n_2-m-2}\int_{F>0}\mathrm{tr}[Q(I+F)F^2](dF).$$

Using Lemma 2.2 and simplifying, we obtain

$$2\int_{F>0}\mathrm{tr}[Q\ (I+F)F^2](dF) = \frac{(n_1+m+1)(\mathrm{tr}\,\Delta)+\mathrm{tr}(\Delta^2)}{n_2-m-1}\,. \tag{2.7}$$

Note that the right hand side of (2.7) is equal to

$$\int_{F>0}\mathrm{tr}(\Delta F)g_1(F)(dF),$$

where $g_1(F)$ is the density of a $F_m(n_1+m+1, n_2; I; \Delta)$ distribution. It follows from (2.7) that

$$\mathrm{tr}[2Q\ (I+F)F^2 g_1^{-1}(F) - \Delta F] = 0 \quad \text{a.e.}$$

or

$$\mathrm{tr}\{[2Q\ g_1^{-1}(F) - F^{-1}\Delta(I+F)^{-1}][(I+F)F^2]\} = 0 \quad \text{a.e.}$$

for all $F > 0$ which implies (2.6). This completes the proof.

3. Improved Estimation of Noncentrality Matrix

The F identity (1.2) is very useful for finding bounds for expectations which often appeares in risk calculations (see for example Konno (1991a) and Leung (1992)). We expect similar applications can be found for the noncentral F identity (2.1) as well. To illustrate a nontrivial application of this noncentral F identity, we consider the problem of estimating the eigenvalues of the noncentrality matrix of a noncentral multivariate F distribution. This problem arises from MANOVA and canonical correlation contexts, and is discussed in Leung and Muirhead (1987) and Leung (1994b).

In the typical MANOVA setting, independent $m \times m$ matrices S_1 and S_2 are observed, where $S_1 \sim W_m(n_1, \Sigma, \Omega)$ and $S_2 \sim W_m(n_2, \Sigma)$. Assume that $n_1 \geq m$ and $n_2 \geq m$, so that both distributions are nonsingular. The eigenvalues of Ω, $\omega_1, \ldots, \omega_m$, are important in the problem of testing $H : \Omega = 0$ against $K : \Omega \neq 0$. Any invariant test depends only on $l_1, \ldots, l_m$, the eigenvalues of $S_1 S_2^{-1}$ and has a power function which depends on Σ and Ω only through $\omega_1, \dots, \omega_m$. These eigenvalues also play a major role in discriminant analysis as well. Now define $m \times m$ matrices A and B by $A = \Sigma^{-1/2} S_1 \Sigma^{-1/2}$ and $B = \Sigma^{-1/2} S_2 \Sigma^{-1/2}$, so that $A \sim W_m(n_1, I, \Delta)$, with $\Delta = \Sigma^{1/2} \Omega \Sigma^{-1/2}$ and $B \sim W_m(n_2, I)$. Therefore $F = A^{1/2} B^{-1} A^{1/2}$ has a $F_m(n_1, n_2; I; \Delta)$ distribution. Note that the eigenvalues of Ω and Δ are the same and the eigenvalues of F and $S_1 S_2^{-1}$ are the same. We remark that, although F is not observable unless Σ is known, its eigenvalues are observable. We then treat F as if it is observable and estimate Δ by $\widehat{\Delta}(F)$ using the invariant loss function

$$L(\Delta, \widehat{\Delta}) = \operatorname{tr}(\Delta^{-1} \widehat{\Delta} - I_m)^2. \tag{3.1}$$

The eigenvalues of $\widehat{\Delta}(F)$ are observable and may be regarded as estimates of $\omega_1, \dots, \omega_m$.

From (i) of Lemma 2.2, the unbiased estimator of Δ is

$$\widehat{\Delta}_U = (n_2 - m - 1)F - n_1 I_m. \tag{3.2}$$

The corresponding estimate of ω_i derived from $\widehat{\Delta}_U$ is thus $(n_2 - m - 1)l_i - n_1$. Now consider two classes of orthogonally invariant estimators,

$$\widehat{\Delta}_\alpha = \alpha \widehat{\Delta}_U \tag{3.3}$$

and

$$\widehat{\Delta}_{\alpha,\beta} = \alpha \widehat{\Delta}_U + \frac{\beta}{\operatorname{tr} F} I_m. \tag{3.4}$$

It is shown in Theorems 3.3 and 3.5 below that $\widehat{\Delta}_\alpha$ dominates $\widehat{\Delta}_U$ for suitable choices of α and $\widehat{\Delta}_{\alpha,\beta}$ dominates $\widehat{\Delta}_\alpha$ for suitable choices of α and β. Before stating and proving the dominance results, we need the following lemmas.

Lemma 3.1.

$$E \operatorname{tr}(\Delta^{-1} F \Delta^{-1} F) = a_1 (\operatorname{tr} \Delta^{-1})^2 + a_2 \operatorname{tr}(\Delta^{-2}) + a_3 \operatorname{tr}(\Delta^{-1}) + a_4,$$

where $a_1 = n_1(n_1 + c_1)/(c_0 c_1 c_3)$, $a_2 = n_1[(n_1 + 1)c_1 + 2]/(c_0 c_1 c_3)$, $a_3 = 2[(n_1 + m + 1)c_1 + m n_1 + 2]/(c_0 c_1 c_3)$, $a_4 = m(m + c_1)/(c_0 c_1 c_3)$ *and* $c_i = n_2 - m - i$.

Proof. The proof is similar to the proof of Lemma 2.1 in Leung (1994b) and is omitted.

Lemma 3.2. *Assume that* $n_2 > m + 3$. *The risk of the unbiased estimator* $\widehat{\Delta}_U$ *in* (3.2) *using loss function* (3.1) *is*

$$R(\Delta, \widehat{\Delta}_U) = b_1 (\operatorname{tr} \Delta^{-1})^2 + b_2 \operatorname{tr}(\Delta^{-2}) + b_3 \operatorname{tr}(\Delta^{-1}) + b_4,$$

where $b_1 = n_1c_1(n_1 + c_1)/(c_0c_3)$, $b_2 = n_1(c_1 + 2)(n_1 + c_1)/(c_0c_3)$, $b_3 = 2(n_1 + c_1)[(m+1)c_1 + 2]/(c_0c_3)$, $b_4 = m[(m+1)c_1 + 2]/(c_0c_3)$ *and* $c_i = n_2 - m - i$.

Proof. The risk of $\widehat{\Delta}_U$ is

$$\begin{aligned} R(\Delta, \widehat{\Delta}_U) &= E\,\mathrm{tr}[\Delta^{-1}\widehat{\Delta}_U - I_m]^2 \\ &= E\,\mathrm{tr}[(n_2 - m - 1)\Delta^{-1}F - n_1\Delta^{-1} - I_m]^2 \\ &= (n_2 - m - 1)^2 E\,\mathrm{tr}(\Delta^{-1}F\Delta^{-1}F) \\ &\quad - 2n_1(n_2 - m - 1)E\,\mathrm{tr}(\Delta^{-2}F) + n_1^2\,\mathrm{tr}(\Delta^{-2}) - m. \end{aligned}$$

Using Lemma 3.1 and the fact that $E\,\mathrm{tr}(\Delta^{-2}F) = [n_1\,\mathrm{tr}(\Delta^{-2}) + \mathrm{tr}(\Delta^{-1})]/c_1$, the result follows after simplification.

Theorem 3.3. *Assume that* $n_2 > m+3$. *Applying the loss function* (3.1), $\alpha\widehat{\Delta}_U$ *dominates* $\widehat{\Delta}_U$ *provided that*

$$\max\left\{0, \frac{c_0c_4 - mc_1 - 1}{c_1(m + c_1)}\right\} < \alpha < 1.$$

Proof. The risk of $\alpha\widehat{\Delta}_U$ is

$$R(\Delta, \alpha\widehat{\Delta}_U) = E[\mathrm{tr}(\alpha\Delta^{-1}\widehat{\Delta}_U - I_m)^2] = \alpha^2 R(\Delta, \widehat{\Delta}_U) + m(1-\alpha)^2.$$

Therefore the difference between the risks of $\widehat{\Delta}_U$ and $\alpha\widehat{\Delta}_U$ is

$$H(\Delta) = R(\Delta, \widehat{\Delta}_U) - R(\Delta, \alpha\widehat{\Delta}_U) = (1-\alpha^2)R(\Delta, \widehat{\Delta}_U) - m(1-\alpha)^2.$$

Using Lemma 3.2,

$$\begin{aligned} H(\Delta) = {} & b_1(1-\alpha^2)(\mathrm{tr}\,\Delta^{-1})^2 + b_2(1-\alpha^2)tr(\Delta^{-2}) + b_3(1-\alpha^2)\,\mathrm{tr}(\Delta^{-1}) \\ & + b_4(1-\alpha^2) - m(1-\alpha)^2. \end{aligned} \tag{3.5}$$

$\alpha\widehat{\Delta}_U$ has a smaller risk than $\widehat{\Delta}_U$ if $H(\Delta) > 0$. However, $H(\Delta)$ depends on the unknown parameter matrix Δ. We need to find a lower bound of $H(\Delta)$ which is independent of Δ. First assume that $0 < \alpha < 1$; then the first three terms in (3.5) are greater than or equal to zero. Therefore

$$H(\Delta) \geq b_4(1-\alpha^2) - m(1-\alpha)^2. \tag{3.6}$$

A sufficient condition for $H(\Delta) > 0$ is $(m - b_4)/(m + b_4) < \alpha < 1$. Note that $(m-b_4)/(m+b_4)$ is always less than 1. The proof is completed after simplification.

An optimal value of α which maximizes the lower bound in (3.6) is

$$\alpha^* = \frac{m}{m + b_4} = \frac{c_0c_3}{c_1(m + c_1)} \tag{3.7}$$

and the corresponding estimate is $\widehat{\Delta}_L = \alpha^* \widehat{\Delta}_U$. Note that α^* always lies between 0 and 1 and satifies the condition in Theorem 3.3.

We now turn to nonlinear estimates $\widehat{\Delta}_{\alpha,\beta}$ defined in (3.4). It is shown in Theorem 3.5 that $\widehat{\Delta}_{\alpha,\beta}$ dominates $\widehat{\Delta}_\alpha$ for suitable choices of α and β. Before stating and proving this dominance result, we need the following lemma.

Lemma 3.4. *Let F have a $F_m(n_1, n_2; I; \Delta)$ distribution with $n_1 > 4$. Then*

$$E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{\operatorname{tr} F}\right] \le \frac{n_1}{n_2-m-1} E\left[\frac{\operatorname{tr}\Delta^{-2}}{\operatorname{tr} F}\right] - \frac{2(n_1-4)}{(n_2-m+3)(n_2-m-1)} E\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr} F)^2}\right] - \frac{2}{(n_2-m-3)(n_2-m+1)} E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{(\operatorname{tr} F)^2}\right] + \frac{1}{n_2-m-1} E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr} F}\right],$$

where E is taken over a $F_m(n_1, n_2; I; \Delta)$ distribution and E_1 is taken over a $F_m(n_1+m+1, n_2; I; \Delta)$ distribution.

Proof. We apply the noncentral F identity given in (2.1) with $V = (I+F)\Delta^{-2}F$ and $h = 1/\operatorname{tr} F$. Since $\operatorname{tr}(DV) = [(m+1)/2](\operatorname{tr}\Delta^{-2}) + (m+1)\operatorname{tr}(\Delta^{-2}F)$ and $\partial h/\partial F = [-1/(\operatorname{tr} F)^2]I_m$ (see Konno (1991a)), we have

$$\frac{n_2-m-1}{n} E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{\operatorname{tr} F}\right] = \frac{n_1}{n} E\left[\frac{\operatorname{tr}\Delta^{-2}}{\operatorname{tr} F}\right] - \frac{2}{n} E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{(\operatorname{tr} F)^2}\right] - \frac{2}{n} E\left[\frac{\operatorname{tr}(\Delta^{-2}F^2)}{(\operatorname{tr} F)^2}\right] + \frac{1}{n} E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr} F}\right]. \quad (3.8)$$

Using the fact that the third term of the right hand side of (3.8) is nonnegative, hence

$$E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{\operatorname{tr} F}\right] \le \frac{n_1}{n_2-m-1} E\left[\frac{\operatorname{tr}\Delta^{-2}}{\operatorname{tr} F}\right] - \frac{2}{n_2-m-1} E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{(\operatorname{tr} F)^2}\right] + \frac{1}{n_2-m-1} E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr} F}\right]. \quad (3.9)$$

To compute the second term of (3.9), apply the noncentral F identity (2.1) again with $V = (I+F)\Delta^{-2}F$ and $h = 1/(\operatorname{tr} F)^2$. Since $\partial h/\partial F = [-2/(\operatorname{tr} F)^3]I_m$ (see Konno (1991a)),

$$\frac{n_2-m-1}{n} E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{(\operatorname{tr} F)^2}\right] = \frac{n_1}{n} E\left[\frac{\operatorname{tr}\Delta}{(\operatorname{tr} F)^2}\right] - \frac{4}{n} E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{(\operatorname{tr} F)^3}\right] - \frac{4}{n} E\left[\frac{\operatorname{tr}(\Delta^{-2}F^2)}{(\operatorname{tr} F)^3}\right] + \frac{1}{n} E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{(\operatorname{tr} F)^2}\right]. \quad (3.10)$$

Using the fact that $\operatorname{tr}(\Delta^{-2}F) \leq (\operatorname{tr}\Delta^{-2})(\operatorname{tr}F)$ and $\operatorname{tr}(\Delta^{-2}F^2) \leq (\operatorname{tr}\Delta^{-2}F)(\operatorname{tr}F)$ in the second and third term of the right hand side of (3.10) respectively, we obtain

$$E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{(\operatorname{tr}F)^2}\right] \geq \frac{n_1-4}{n_2-m+3}E\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr}F)^2}\right] + \frac{1}{n_2-m+3}E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{(\operatorname{tr}F)^2}\right].$$

Substituting into (3.9) completes the proof.

Theorem 3.5. *Assume that* $n_1 > 4$ *and* $n_2 > m-1$. *Then* $\widehat{\Delta}_{\alpha,\beta}$ *defined in* (3.4) *dominates* $\widehat{\Delta}_\alpha$ *defined in* (3.3) *if*

$$0 < \alpha < 1 + \frac{2}{m(n_2-m+1)} \quad \text{and} \quad 0 < \beta < \frac{4\alpha(n_1-4)}{(n_2-m+3)}.$$

Proof. For the loss function defined in (3.1), it is straight forward to show that the risk difference between $\widehat{\Delta}_\alpha$ and $\widehat{\Delta}_{\alpha,\beta}$ is

$$\begin{aligned} G(\Delta) &= E[L(\Delta,\widehat{\Delta}_\alpha) - L(\Delta,\widehat{\Delta}_{\alpha,\beta})] \\ &= 2\beta E\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr}F}\right] - 2(n_2-m-1)\alpha\beta E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{\operatorname{tr}F}\right] + 2n_1\alpha\beta E\left[\frac{\operatorname{tr}\Delta^{-2}}{\operatorname{tr}F}\right] \\ &\quad - \beta^2 E\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr}F)^2}\right]. \end{aligned}$$

Using Lemma 3.4 and simplifying, we obtain

$$\begin{aligned} G(\Delta) \geq 2\beta\left\{E\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr}F}\right] - \alpha E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr}F}\right]\right\} + \frac{4\alpha\beta}{(n_2-m+3)}E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{(\operatorname{tr}F)^2}\right] \\ + \beta\left\{\frac{4\alpha(n_1-4)}{(n_2-m+3)} - \beta\right\}E\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr}F)^2}\right]. \end{aligned} \tag{3.11}$$

First, assume that α and β are positive. Then the second and the third terms on the right hand side of (3.11) are always positive. The first term involves $E[1/\operatorname{tr}F]$. An upper and lower bound for $E[1/\operatorname{tr}F]$ is given in Lemma 3.4 in Leung and Muirhead (1987) (derived using the Wishart identity) as follow :

$$\frac{1}{m}E\left[\frac{2+m(n_2-m-1)}{mn_1+2K-2}\right] \leq E\left[\frac{1}{\operatorname{tr}F}\right] \leq E\left[\frac{n_2-m+1}{mn_1+2K-2}\right],$$

where K is a Poisson random variable with mean $(\operatorname{tr}\Delta)/2$. Using the lower bound for $E[1/\operatorname{tr}F]$ and the upper bound for $E_1[1/\operatorname{tr}F]$ and simplifying, the first term on the right hand side of (3.11) is apparently equal to

$$2\beta\operatorname{tr}(\Delta^{-1})E\left\{\frac{(mn_1+2K-2)[(1-\alpha)m(n_2-m+1)+2]+2m(m+1)}{m(mn_1+2K-2)[m(n_1+m+1)+2K-2]}\right\}$$

which is greater than zero if the square bracket in the numerator is greater than zero. This is exactly the condition of α stated in the Theorem. Therefore

$$G(\Delta) \geq \beta \left\{ \frac{4\alpha(n_1-4)}{(n_2-m+3)} - \beta \right\} E\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr} F)^2} \right] \tag{3.12}$$

and a sufficient condition for $G(\Delta) \geq 0$ is to ensure the curly bracket of (3.12) is nonnegative. This completes the proof.

Table 1. — PRIAL : $\Delta = \operatorname{diag}(1,1,1,1)$

n_1	n_2	$\widehat{\Delta}_U^+$	$\widehat{\Delta}_L$	$\widehat{\Delta}_L^+$	$\widehat{\Delta}_{NL}$	$\widehat{\Delta}_{NL}^+$
10	10	10.341	83.889	86.077	83.935	86.038
	25	28.146	37.946	56.494	38.292	55.546
	50	29.876	19.550	44.251	19.894	42.657
	75	33.009	13.135	42.312	13.461	40.316
	100	32.357	9.893	39.429	10.222	37.325
25	10	21.418	83.901	87.878	83.962	87.816
	25	27.021	38.017	55.360	38.152	54.615
	50	34.625	19.534	47.842	19.718	46.336
	75	37.588	13.128	46.124	13.320	44.224
	100	39.479	9.871	45.725	10.198	43.670
50	10	15.064	83.988	86.704	84.010	86.675
	25	28.060	37.993	55.753	38.077	55.268
	50	37.412	19.523	49.933	19.668	48.841
	75	39.712	13.119	47.848	13.303	46.516
	100	40.851	9.875	46.879	10.103	45.361
75	10	13.830	83.992	86.339	84.005	86.321
	25	26.659	37.992	56.655	38.046	56.277
	50	34.971	19.517	47.886	15.595	47.095
	75	38.437	13.133	46.691	13.200	45.651
	100	41.215	9.886	47.171	9.983	45.944
100	10	16.599	83.997	86.775	84.007	86.758
	25	29.435	37.993	56.456	38.031	56.158
	50	36.832	19.532	49.344	19.582	48.684
	75	36.791	13.136	45.228	13.163	44.384
	100	41.314	9.882	47.229	10.000	46.254

A reasonable way of choosing β is by maximizing the lower bound for $G(\Delta)$ in (3.12). The maximizing value is $\beta^* = 2\alpha^*(n_1-4)/(n_2-m+3)$ where α^* is defined in (3.7) and the corresponding estimator is $\widehat{\Delta}_{NL} = \widehat{\Delta}_{\alpha^*,\beta^*}$.

$\widehat{\Delta}_U$, $\widehat{\Delta}_L$ and $\widehat{\Delta}_{NL}$ are not necessarily positive definite; they are dominated by their truncated versions $\widehat{\Delta}_U^+$, $\widehat{\Delta}_L^+$ and $\widehat{\Delta}_{NL}^+$ respectively: matrices with the same eigenvectors and eigenvalues except that any negative eigenvalues are replaced by zero. This result is proved in Theorem 3.3 in Leung (1994a).

Table 2. — PRIAL : $\Delta = \text{diag}(4,3,2,1)$

n_1	n_2	$\widehat{\Delta}_U^+$	$\widehat{\Delta}_L$	$\widehat{\Delta}_L^+$	$\widehat{\Delta}_{NL}$	$\widehat{\Delta}_{NL}^+$
10	10	12.405	83.647	86.698	83.732	86.653
	25	20.339	37.959	51.959	38.226	51.245
	50	26.470	19.494	41.785	19.971	40.476
	75	28.797	13.094	38.858	13.556	37.193
	100	27.272	9.894	35.028	10.205	33.178
25	10	16.212	83.858	87.264	83.921	87.211
	25	26.276	22.766	55.267	38.171	54.521
	50	30.539	19.544	44.850	19.703	43.290
	75	31.960	13.144	41.434	13.335	39.592
	100	32.475	9.894	39.583	10.035	37.531
50	10	17.503	83.932	87.212	83.966	87.177
	25	27.831	37.942	55.876	38.125	55.365
	50	31.348	19.532	45.239	19.651	44.162
	75	23.991	13.133	42.648	13.252	41.273
	100	37.314	18.077	43.810	10.152	42.275
75	10	9.660	84.002	85.740	84.015	85.725
	25	25.063	37.996	54.016	38.053	53.608
	50	33.872	19.514	47.150	19.670	46.314
	75	35.330	13.122	44.119	13.261	42.996
	100	37.480	9.881	43.903	10.025	42.625
100	10	15.157	83.982	86.629	83.997	86.611
	25	31.606	37.929	57.748	38.059	57.639
	50	27.441	19.555	41.907	19.551	41.227
	75	35.746	13.129	44.433	13.226	43.498
	100	37.200	9.883	43.606	10.006	42.569

A simulation study was carried out to compare the risks of $\widehat{\Delta}_U$, $\widehat{\Delta}_L$ and $\widehat{\Delta}_{NL}$ and their truncated versions $\widehat{\Delta}_U^+$, $\widehat{\Delta}_L^+$ and $\widehat{\Delta}_{NL}^+$. For $m = 4$ and $n_1, n_2 = 10, 25, 50, 75, 100$, and three different choices of a diagonal noncentrality matrix Δ, a sample of 500 A's and B's were generated, where $A \sim W_4(n_1, I, \Delta)$, $B \sim W_4(n_2, I)$, A and B are independent. Then 500 values of $F = A^{1/2}B^{-1}A^{-1/2}$ were formed and used to construct $\widehat{\Delta}_U$, $\widehat{\Delta}_L$, $\widehat{\Delta}_{NL}$ and their truncated versions $\widehat{\Delta}_U^+$, $\widehat{\Delta}_L^+$ and $\widehat{\Delta}_{NL}^+$, and from these average losses were obtained. Tables 1 to 3 give the percentage reduction in average loss (PRIAL) for $\widehat{\Delta}_U^+$, $\widehat{\Delta}_L$, $\widehat{\Delta}_L^+$, $\widehat{\Delta}_{NL}$ and $\widehat{\Delta}_{NL}^+$

compared with $\widehat{\Delta}_U$, i.e., they are the estimates of

$$\frac{R(\Delta, \widehat{\Delta}_U) - R(\Delta, \widehat{\Delta})}{R(\Delta, \widehat{\Delta}_U)} \times 100$$

obtained by replacing risk with average loss and $\widehat{\Delta}$ with various estimators.

PRIAL in Tables 1 to 3 are all positive, which confirms the dominance results. The risk reduction of $\widehat{\Delta}_U^+$ is small when n_2 is small. $\widehat{\Delta}_L^+$ and $\widehat{\Delta}_{NL}^+$ are uniformly better than $\widehat{\Delta}_U^+$. $\widehat{\Delta}_L$ and $\widehat{\Delta}_{NL}$ and their truncated version are substantially better than $\widehat{\Delta}_U^+$ when n_2 is small.

Table 3. — PRIAL : $\Delta = \mathrm{diag}(100, 75, 50, 25)$

n_1	n_2	$\widehat{\Delta}_U^+$	$\widehat{\Delta}_L$	$\widehat{\Delta}_L^+$	$\widehat{\Delta}_{NL}$	$\widehat{\Delta}_{NL}^+$
10	10	0.482	71.656	71.862	71.666	71.870
	25	0.314	30.089	30.197	30.138	30.243
	50	0.223	17.224	17.426	17.267	17.462
	75	0.064	11.682	11.741	11.722	11.780
	100	0.113	8.831	8.938	8.869	8.974
25	10	3.298	74.664	76.002	74.687	76.013
	25	2.464	33.207	35.159	33.303	35.206
	50	1.267	17.832	18.978	17.928	19.035
	75	0.755	11.701	12.407	11.812	12.484
	100	0.992	8.945	9.886	9.046	9.949
50	10	7.023	78.491	81.160	78.516	81.165
	25	7.728	34.499	40.532	34.611	40.525
	50	7.116	17.983	24.341	18.121	24.311
	75	4.568	12.690	16.946	12.764	16.884
	100	4.456	9.349	13.580	9.461	13.535
75	10	10.018	80.139	83.688	80.162	83.688
	25	13.150	35.926	46.024	36.019	45.967
	50	11.065	18.615	28.448	18.735	28.345
	75	10.077	12.260	21.600	12.394	21.447
	100	8.517	9.551	17.610	9.650	17.459
100	10	12.468	80.731	84.843	80.753	84.842
	25	18.052	35.858	49.501	35.966	49.431
	50	12.779	19.433	30.755	19.477	30.577
	75	13.181	12.670	24.855	12.767	24.662
	100	12.887	9.537	21.705	9.643	21.492

Acknowledgement

The authors wish to thank two referees for their helpful comments. The research of the first author was supported by Direct Grant of the Chinese University of Hong Kong.

References

Haff, L. R. (1979a). An identity for the Wishart distribution with applications. J. Multivariate Anal. **9,** 531–544.

Haff, L. R. (1979b). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. Ann. Statist. **7,** 1264–1276.

Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. Ann. Statist. **8,** 586–597.

Haff, L. R. (1981). Further identities for the Wishart distribution with applications in regression. Candian. J. Statist. **9,** 215–224.

Haff, L. R. (1982). Identities for the inverse Wishart distribution with computational results in linear and quadratic discrimination. Sankhyā Ser.B **44,** 245–258.

Konno, Y. (1988). Exact moments of the multivariate F and Beta distributions. J. Japan Statist. Soc. **18**, 123–130.

Konno, Y. (1991a). A note on estimating eigenvalues of scale matrix of the multivariate F-distribution. Ann. Inst. Statist. Math. **43**, 157–165.

Konno, Y. (1991b). On Estimation of a Matrix of Normal Means with Unknown Covariance Matrix. J. Multivariate Anal. **36**, 44–55.

Konno, Y. (1992). On estimating eigenvalues of the scale matrix of the multivariate F distribution. Sankhyā Ser.A **54,** 241–251.

Leung, P. L. (1986). Estimation of parameter matrices and their eigenvalues in multivariate analysis. Unpublished Ph.D. thesis. University of Michigan.

Leung, P. L. (1992). Estimation of eigenvalues of the scale matrix of the multivariate F distribution. Comm. Statist. Theory Methods **21**, 1845–1856.

Leung, P. L. (1994a). An identity for the noncentral Wishart distribution with application. J. Multivariate Anal. **48**, 107–114.

Leung, P. L. (1994b). Improved estimation of functions of canonical correlation coefficients. Comm. Statist. Theory Methods **23**, 831–840.

Leung, P. L. and Muirhead, R. J. (1987). Estimation of parameter matrices and eigenvalues in MANOVA and canonical correlation analysis. Ann. Statist. **15**, 1651–1666.

Magnus, J. R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications. Ann. Statist. **7**, 381–394.

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

Muirhead, R. J. and Verathaworn, T. (1985). On estimating the latent roots of $\Sigma_1\Sigma_2^{-1}$. In Multivariate Analysis VI (Edited by P. R. Krishnaiah), 431-447, North-Holland, Amsterdam.

Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong.

(Received March 1993; accepted March 1995)