Ann. Inst. Statist. Math.
Vol. 53, No. 4, 769–780 (2001)

IMPROVED ESTIMATION OF PARAMETER MATRICES IN A ONE-SAMPLE AND TWO-SAMPLE PROBLEMS

PUI LAM LEUNG AND FOON YIP NG

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

(Received October 19, 1999; revised July 14, 2000)

Abstract. In this paper, the problems of estimating the covariance matrix in a Wishart distribution (refer as one-sample problem) and the scale matrix in a multivariate F distribution (which arise naturally from a two-sample setting) are considered. A new class of estimators which shrink the eigenvalues towards their harmonic mean is proposed. It is shown that the new estimator dominates the best linear estimator under two scale invariant loss functions.

Key words and phrases: Covariance matrix, orthogonally invariant estimator, decision–theoretic estimation, shrinkage estimator, harmonic mean, eigenvalues.

1. Introduction and summary

Suppose that a random $m \times m$ positive definite matrix S has a nonsingular Wishart distribution with unknown covariance matrix Σ and n degrees of freedom, i.e., $S \sim W_m(n, \Sigma)$. There has been considerable research in estimating the covariance matrix Σ using a decision–theoretic approach. It is shown that substantial improvement (reduction in risks) over the usual unbiased estimator of Σ can be obtained, essentially by focusing attention on the problem of estimating the eigenvalues of Σ by functions of all the eigenvalues of Σ. In particular, Stein (1975) considered the class of orthogonally invariant estimators of Σ of the form

$$\hat{\Sigma} = H\Phi(Q)H' \tag{1.1}$$

where $S = HQH'$ with H the matrix of normalized eigenvectors ($HH' = H'H = I_m$), $Q = \mathrm{diag}(l_1, \ldots, l_m)$ is the diagonal matrix of eigenvalues of S with $l_1 > \cdots > l_m > 0$ and $\Phi(Q) = \mathrm{diag}(\phi_1(Q), \ldots, \phi_m(Q))$, $\phi_i(Q) \geq 0$ is a real valued function, $i = 1, \ldots, m$. Works along this direction can be found in Haff (1980), Dey and Srinivasan (1985), Lin and Perlman (1985) and Dey (1988). Excellent reviews on this topic can be found in Muirhead (1987) and Pal (1993). However, all the proposed estimators will shrink the eigenvalues towards the origin. In practice, it is more useful and reasonable to shrink the eigenvalues toward some central values.

In this paper, we consider the problem of estimating Σ using two scale–invariant loss functions

$$L_1(\Sigma, \hat{\Sigma}) = \mathrm{tr}(\hat{\Sigma}\Sigma^{-1}) - \ln|\hat{\Sigma}\Sigma^{-1}| - m, \tag{1.2}$$

and

$$L_2(\Sigma, \hat{\Sigma}) = \operatorname{tr}(\hat{\Sigma}\Sigma^{-1} - I_m)^2. \tag{1.3}$$

Haff (1980) shows that for $c_1 = 1/n$ and $c_2 = 1/(n+m+1)$, c_1S and c_2S has the smallest risk among all the scalar multiples of S under L_1 and L_2 loss function respectively. We refer these estimators as the best linear estimators thereafter. We proposed a new estimator of the form

$$\hat{\Sigma}(\alpha) = \alpha cS + (1-\alpha)c\frac{m}{\operatorname{tr} S^{-1}} I_m \tag{1.4}$$

where $0 \le \alpha \le 1$. c is a constant appropriately chosen for the two loss functions (1.2) and (1.3), i.e., c_1 for L_1 loss function and c_2 for L_2 loss function. We prove that c_1S and c_2S are inadmissible and is dominated by $\hat{\Sigma}(\alpha)$ in (1.4) for some values of α. Note that for $\alpha = 1$, $\hat{\Sigma}(\alpha)$ corresponds to the linear estimator cS. $\hat{\Sigma}(\alpha)$ is in the class of orthogonally invariant estimators defined in (1.1) with eigenvalues

$$\begin{aligned} \phi_i(L) &= \alpha c l_i + (1-\alpha)c\tilde{l}, \\ &= c l_i - (1-\alpha)c(l_i - \tilde{l}) \qquad (i = 1, \ldots, m) \end{aligned} \tag{1.5}$$

where $\tilde{l} = m/\operatorname{tr} S^{-1}$ is the harmonic mean of $l_1, \ldots, l_m$, the eigenvalues of S.

This estimator is motivated by the fact that the sample eigenvalues of S usually tend to be much more dispersed than the population eigenvalues of Σ (see Muirhead (1987)), i.e., l_1 tends to over–estimate the largest population eigenvalue while l_m tends to under–estimate the smallest population eigenvalue. Intuitively, cS can be improved by shrinking the sample eigenvalues towards some central value. From equation (1.5), it is easy to see that $\phi_i(L)$ is obtained by shrinking cl_i, the eigenvalues of cS, towards their harmonic mean $c\tilde{l}$. The amount of shrinkage depends on the parameter α and the distance between l_i and $\tilde{l}$. α is the shrinkage parameter ranging from 0 to 1 representing various degrees of shrinkage. Note also that $\hat{\Sigma}(\alpha)$ in (1.4) preserves the order of the sample eigenvalues l_i. This class of order-preserving othogonally invariant estimator is well studied by Sheena and Takemura (1992) as it form a complete class.

This idea of shrinking the eigenvalues towards certain central value is not new. It has been used in Friedman (1989) and Leung and Chan (1998) where the arithmetic mean, instead of the harmonic mean, was adopted as the central value. In Friedman (1989), he used this shrinkage in discriminant analysis, and coined the name 'regularized discriminant analysis'. In Leung and Chan (1998), they considered the same estimation problem using the squared error loss function. However, this squared error loss function is not scale invariant. We would like to develop similar dominance results for the L_1 and L_2 loss functions. If we shrink the eigenvalues towards their sample mean as in Leung and Chan (1998), it is extremely difficult to obtain dominance results. If we use the harmonic mean as the central value, we are able to obtain dominance results as in Sections 2 and 3. Another possible way is to shrink the eigenvalues towards their geometric mean, but this possibility will not be explored in the paper and leave for further study.

Another closely related problem is the estimation of the scale matrix Δ in a multivariate F distribution. This problem has been considered by various authors, namely, Muirhead and Verathaworn (1985), Leung and Muirhead (1988), Dey (1989), Gupta and

Krishnamoorthy (1990), Konno (1991) and Leung (1992). This problem arises naturally from a two-sample setting (see Muirhead and Verathaworn (1985) and Leung and Muirhead (1988)).

Suppose that a random $m \times m$ positive definite matrix F has a multivariate F distribution with degrees of freedom n_1 and n_2 and scale matrix Δ, i.e., $F \sim F_m(n_1, n_2; \Delta)$. We consider the two invariant loss functions (1.2) and (1.3) (with Σ replaced by Δ). Muirhead and Verathaworn (1985) show that $c_3 F$, where

$$c_3 = (n_2 - m - 1)/n_1, \tag{1.6}$$

is the best among all scalar multiple of F under the L_1 loss function when $n_2 > m + 1$. Later on, Leung and Muirhead (1988) show that $c_4 F$, where

$$c_4 = \frac{(n_2 - m)(n_2 - m - 3)}{(n_2 - m - 1)(n_1 + m + 1) + mn_1 + 2}, \tag{1.7}$$

is the best among all scalar multiples of F under the L_2 loss function when $n_2 > m + 3$. Using a similar approach as in the Wishart situation, we propose a new class of orthogonally invariant estimators of the form

$$\hat{\Delta}(\alpha) = \alpha c F + (1 - \alpha) c \frac{m}{\operatorname{tr} F^{-1}} I_m \tag{1.8}$$

where $0 \leq \alpha \leq 1$.

The present paper is organized as follows: In Section 2, we considered the problem of estimating the covariance matrix Σ in a Wishart distribution (one–sample problem) and the scale matrix Δ in a multivariate F distribution (two–sample problem) under the L_1 loss function. We proved that the best linear estimator $c_1 S$ (and $c_3 F$) is inadmissible and is dominated by $\hat{\Sigma}(\alpha)$ in (1.4) (and $\hat{\Delta}(\alpha)$ in (1.8)) for some suitable chosen values of α. Similarly in Section 3, we considered the same estimation problem using the L_2 loss function. Finally, a simulation study is carried out to study the performance of these new estimators in Section 4.

In Sections 2 and 3, some expectations involved in the risk calculations are very complicated and cannot be expressed in closed forms. However, in order to prove the dominance results, only upper and lower bounds for these expectations are needed. These bounds are formulated as the following lemma and will be used in Sections 2 and 3.

LEMMA 1. *Assume that $m > 1$, $n_1 > m + 1$ and $n_2 > m + 1$. Let $F \sim F_m(n_1, n_2; \Delta)$.*

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

(ii) $E[\frac{\operatorname{tr}(\Delta^{-2} F)}{\operatorname{tr} F^{-1}}] \leq [\frac{(n_1 + 2)(n_2 - 2) - 2(m - 1)}{(n_2 - m - 1)(n_2 - 4)}] E[\frac{\operatorname{tr} \Delta^{-1}}{\operatorname{tr} F^{-1}}]$,

(iii) $\frac{n_1 - m - 1}{n_2} \leq E[\frac{\operatorname{tr} \Delta^{-1}}{\operatorname{tr} F^{-1}}] \leq \frac{n_1 - m + 1}{n_2 - 2}$.

Lemma 1 can be proved by the multivariate F identity (see Muirhead and Verathaworn (1985) and Konno (1988)). The proof of this lemma together with the multivariate F identities are given in the Appendix.

2. Improved estimation of Σ and Δ under L_1 loss function

First we consider the problem of estimating the scale matrix Δ in a multivariate F distribution. Estimation of Σ in a Wishart distribution can be considered as a limiting case of the former problem and will be considered later. Let $A \sim W_m(n_1, \Delta)$ and independent of $B \sim W_m(n_2, I)$. Then the random $m \times m$ positive definite matrix $F = A^{1/2}B^{-1}A^{1/2}$ has a multivariate F distribution with degrees of freedom n_1 and n_2 and scale matrix Δ, i.e., $F \sim F_m(n_1, n_2; \Delta)$. Muirhead and Verathaworn (1985) show that c_3F, where c_3 is given in (1.6), is the best linear estimator of Δ. The main result in this section is to provide a sufficient condition on α such that c_3F is dominated by $\hat{\Delta}(\alpha)$ defined in (1.8).

THEOREM 2.1. *Assume that* $m > 1$, $n_1 > m + 1$ *and* $n_2 > m + 1$. *Applying the loss function* (1.2),

$$\hat{\Delta}_1(\alpha) = \alpha c_3 F + (1-\alpha)c_3 \frac{m}{\operatorname{tr} F^{-1}} I_m \tag{2.1}$$

dominates the best linear estimator c_3F *provided that*

$$\frac{-q + \sqrt{q^2 + 8q(m-1)(n_1+n_2-m-1)}}{4(m-1)(n_1+n_2-m-1)} \le \alpha < 1, \tag{2.2}$$

where $q = mn_1(n_2 - 2)$.

PROOF. For the loss function in (1.2), it is straightforward to show that the risk difference between c_3F and $\hat{\Delta}_1(\alpha)$ is

$$\begin{aligned} G_1(\Delta) &= E[L_1(\Delta, c_3F)] - E[L_1(\Delta, \hat{\Delta}_1(\alpha))] \\ &= (1-\alpha)m\left\{1 - c_3 E\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr} F^{-1}}\right]\right\} + E\left[\ln\left|I_m + \frac{(1-\alpha)m}{\alpha(\operatorname{tr} F^{-1})}F^{-1}\right|\right] \\ &\quad + m\ln\alpha. \end{aligned}$$

Using the upper bound in (iii) of Lemma 1 and $\ln|I + A| \ge \operatorname{tr} A - (1/2)\operatorname{tr} A^2$ for any positive definite matrix A (see Konno (1991), p. 160), we have

$$\begin{aligned} G_1(\Delta) \ge{}& \frac{(1-\alpha)m(m-1)(n_1+n_2-m-1)}{n_1(n_2-2)} + m\ln\alpha + \frac{(1-\alpha)m}{\alpha} \\ & - \frac{(1-\alpha)^2m^2}{2\alpha^2}E\left[\frac{\operatorname{tr}(F^{-2})}{(\operatorname{tr} F^{-1})^2}\right]. \end{aligned}$$

Note that $\ln\alpha \ge -(1-\alpha)/\alpha$ and $\operatorname{tr}(F^{-2})/(\operatorname{tr} F^{-1})^2 \le 1$,

$$G_1(\Delta) \ge \frac{(1-\alpha)m}{2n_1(n_2-2)\alpha^2}[2(m-1)(n_1+n_2-m-1)\alpha^2 + q\alpha - q], \tag{2.3}$$

where $q = mn_1(n_2 - 2)$. Note that the term in the square bracket in (2.3) is a quadratic function in α with the coefficient of α^2 being positive. The condition (2.2) in Theorem 2.1 ensures that the lower bound for $G_1(\Delta)$ in (2.3) is non-negative and the proof is completed.

Remark 1. Intuitively speaking, we would like to find a value of α that maximizes the risk difference $G_1(\Delta)$. However, this α generally depends on the unknown matrix Δ. Instead, we suggest using an 'optimal' value of α that maximizes the lower bound in (2.3). It is easy to show that such optimal value is the real root of the cubic equation $\alpha^3 + r\alpha - r = 0$, where

$$r = \frac{mn_1(n_2 - 2)}{(m-1)(n_1 + n_2 - m - 1)}. \tag{2.4}$$

Cartan's solution of this cubic equation (see Merritt (1962), p. 49) is:

$$\alpha_1 = \sqrt[3]{\frac{r}{2} + \sqrt{\left(\frac{r^2}{4} + \frac{r^3}{27}\right)}} + \sqrt[3]{\frac{r}{2} - \sqrt{\left(\frac{r^2}{4} + \frac{r^3}{27}\right)}}. \tag{2.5}$$

Remark 2. An anonymous referee pointed out that the result in Theorem 2.1 can be generalized easily by considering a new class of estimators

$$\hat{\Delta}_N(\alpha) = \alpha c_3 F + (1-\alpha) c_3 \vartheta g(\vartheta) I_m \tag{2.6}$$

where $\vartheta = m/\operatorname{tr} F^{-1}$, and $g(\vartheta)$ is a bounded positive real valued function of ϑ such that $0 < k_1 \le g(\vartheta) \le k_2 < \infty$, for some constants $0 < k_1 \le k_2$. With this $\hat{\Delta}_N(\alpha)$ and similar calculation, one can write the risk difference

$$\begin{aligned} G_N(\Delta) &= E[L_1(\Delta, c_3F)] - E[L_1(\Delta, \hat{\Delta}_N(\alpha))] \\ &\ge \frac{(1-\alpha)m}{\alpha^2}\left\{\left[1 - c_3 k_1\left(\frac{n_1 - m + 1}{n_2 - 2}\right)\right]\alpha^2 + (k_1 - 1)\alpha - \frac{1-\alpha}{2} m k_2^2\right\}. \end{aligned}$$

The expression in the curly bracket is a quadratic function in α. Therefore, one can establish a similar sufficient condition on α such that $G_N(\Delta) > 0$. However, this condition is rather messy and depends on k_1 and k_2. When $k_1 = k_2 \equiv 1$, $\hat{\Delta}_N(\alpha)$ reduces to $\hat{\Delta}_1(\alpha)$ and $G_N(\Delta)$ reduces to $G_1(\Delta)$.

Now we turn to the problem of estimating Σ in a Wishart distribution.

THEOREM 2.2. *Assume that $m > 1$ and $n > m + 1$. Applying the loss function* (1.2),

$$\hat{\Sigma}_1(\alpha) = \alpha c_1 S + (1-\alpha) c_1 \frac{m}{\operatorname{tr} S^{-1}} I_m \tag{2.7}$$

dominates the best linear estimator $c_1 S$ (*where* $c_1 = 1/n$) *provided that*

$$\frac{-nm + \sqrt{n^2m^2 + 8nm(m-1)}}{4(m-1)} \le \alpha < 1. \tag{2.8}$$

PROOF. Note that in the two-sample situation, $F = A^{1/2}B^{-1}A^{1/2}$ where $A \sim W_m(n_1, \Delta)$ and independent of $B \sim W_m(n_2, I)$. B can be considered as the sum of squares and cross product matrix of n_2 independently and identically distributed $m \times 1$ standard normal random vectors. By the strong law of large numbers, $n_2^{-1}B$ converges to I_m or n_2F converges to A almost surely, as n_2 tends to infinity. This becomes the problem of estimating the covariance matrix Σ in a Wishart distribution. Furthermore, n_2F is uniformly integrable so exchange between $\lim_{n_2\to\infty}$ and expectation is possible. Therefore, the sufficient condition (2.8) in Theorem 2.2 can be obtained by letting $n_2 \to \infty$ in the sufficient condition (2.2) in Theorem 2.1 (with n_1 replaced by n) and the proof is completed.

Remark 3. It is possible to prove Theorem 2.2 by computing the risk difference directly as in Theorem 2.1. However, it is easier to prove the result by letting $n_2 \to \infty$. This technique has also been used in Leung and Chan (1998). The 'optimal' value for α (maximizes the risk difference between c_1S and $\hat{\Sigma}(\alpha)$) in this one-sample situation is the same as in equation (2.5) except that $r = mn/(m-1)$. Again this is obtained by letting $n_2 \to \infty$ and n_1 is replaced by n in (2.4).

3. Improved estimation of Σ and Δ under L_2 loss function

Similar to Section 2, we first consider the problem of estimating Δ in a multivariate F distribution. The corresponding result for estimating Σ in a Wishart distribution can be obtained by letting $n_2 \to \infty$. Leung and Muirhead (1988) show that c_4F, where c_4 is given in (1.7), is the best linear estimator of Δ. The main result in this section is to provide a sufficient condition on α such that c_4F is dominated by $\hat{\Delta}(\alpha)$ defined in (1.8).

THEOREM 3.1. *Assume that* $m > 3$,

$$n_1 \ge m + 3 + 12/(m-3), \tag{3.1}$$

$$\begin{aligned} n_2 \ge\ & \frac{\sqrt{[t_1(m+5) - t_2(m+2)]^2 - 8(t_1 - t_2)[2t_1(m+1) - mt_2]}}{2(t_1 - t_2)} \\ & + \frac{t_1(m+5) - t_2(m+2)}{2(t_1 - t_2)}, \end{aligned} \tag{3.2}$$

where $t_1 = (n_1 - m - 1)(n_1 + m + 1)$, $t_2 = (n_1 - m + 1)(n_1 + 2)$ *and*

$$n_2 \ge \frac{m(m+1)(n_1 - m + 3) - 4(m-1)}{(m-1)(n_1 - m + 2)}. \tag{3.3}$$

Applying the loss function (1.3),

$$\hat{\Delta}_2(\alpha) = \alpha c_4 F + (1-\alpha)c_4 \frac{m}{\operatorname{tr} F^{-1}} I_m \tag{3.4}$$

dominates the best linear estimator c_4F *provided that* $b_1/a_1 \le \alpha < 1$ *where*

$$a_1 = k - \frac{2c_4(n_1-m+1)[(n_1+2)(n_2-2)-2(m-1)]}{(n_2-2)(n_2-4)(n_2-m-1)}, \tag{3.5}$$

$$b_1 = k - 2(n_1-m-1)/n_2, \tag{3.6}$$

$$k = \frac{n_1}{n_2-m-1} + \frac{mc_4(n_1-m+1)(n_1-m+3)}{(n_2-2)(n_2-4)}.$$

PROOF. For the loss function in (1.3), it is straightforward to show that the risk difference between c_4F and $\hat{\Delta}_2(\alpha)$ is

$$\begin{aligned} G_2(\Delta) &= E[L_2(\Delta, c_4F)] - E[L_2(\Delta, \hat{\Delta}_2(\alpha))] \\ &= (1-\alpha)mc_4 \left\{ \frac{(1+\alpha)c_4}{m} E[\operatorname{tr}(\Delta^{-1}F)^2] - \frac{2n_1}{n_2-m-1} + 2E\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr}F^{-1}}\right] \right. \\ &\qquad \left. -(1-\alpha)E\left[\frac{mc_4 \operatorname{tr}\Delta^{-2}}{(\operatorname{tr}F^{-1})^2}\right] - 2\alpha c_4 E\left[\frac{\operatorname{tr}(\Delta^{-2}F)}{\operatorname{tr}F^{-1}}\right] \right\}. \end{aligned}$$

Using Corollary 2.4 in Konno (1988), it can be shown that $E[\operatorname{tr}(\Delta^{-1}F)^2] = mn_1/[c_4(n_2-m-1)]$. Applying the upper bounds (i) and (ii) in Lemma 1,

$$\begin{aligned} G_2(\Delta) \ge (1-\alpha)mc_4 &\left\{ \frac{(1+\alpha)n_1 - 2n_1}{n_2-m-1} + 2E\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr}F^{-1}}\right] \right. \\ &- \frac{(1-\alpha)mc_4(n_1-m+3)}{n_2-4} E\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr}F^{-1}}\right] \\ &\left. - \frac{2c_4\alpha[(n_1+2)(n_2-2)-2(m-1)]}{(n_2-m-1)(n_2-4)} E\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr}F^{-1}}\right] \right\}. \end{aligned} \tag{3.7}$$

Note that there are three expectation terms in (3.7). The coefficient of the first expectation is positive while the last two are negative. Therefore we apply the lower bound of (iii) in Lemma 1 to the first expectation and the upper bound of (iii) to the last two expectation terms and simplify,

$$G_2(\Delta) \ge (1-\alpha)mc_4(a_1\alpha - b_1), \tag{3.8}$$

where a_1 and b_1 are given in (3.5) and (3.6). A sufficient condition for $G_2(\Delta) \ge 0$ is $b_1/a_1 \le \alpha < 1$. Conditions (3.1) and (3.2) are needed to ensure $b_1/a_1 < 1$. Condition (3.3) is to ensure $a_1 \ge 0$ and the proof is completed.

Remark 4. It is easy to see that the 'optimal' value for α that maximizes the lower bound in (3.7) is $\alpha_2 = (a_1+b_1)/(2a_1)$ where a_1 and b_1 is given in (3.5) and (3.6) respectively.

Table 1. Minimum integer value of n_2 satisfy conditions (3.3) and (3.4).

n_1	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	$m=20$
20	1128	94	53	40	33	30	28	32
30	241	84	53	40	34	30	27	27
40	228	93	61	46	38	33	30	26
50	242	106	69	53	44	38	34	27
60	264	119	79	60	49	42	38	28
70	289	133	88	67	55	47	42	30
80	316	147	98	74	61	52	46	31
90	244	162	108	82	67	57	50	32
100	372	177	118	89	73	62	54	34

Remark 5. The conditions (3.2) and (3.3) in Theorem 3.1 seem rather restrictive. Table 1 gives the minimum integer value of n_2 that satisfy (3.2) and (3.3) for various combinations of m and n_1. From Table 1, the condition on n_2 becomes less and less restrictive when m increase.

For the problem of estimating Σ in a Wishart distribution, similar results can be obtained from Theorem 3.1 by letting $n_2 \to \infty$ and replace n_1 by n.

THEOREM 3.2. *Assume that $m > 3$ and $n > m + 3 + 12/(m-3)$. Applying the loss function* (1.3),

$$\hat{\Sigma}_2(\alpha) = \alpha c_2 S + (1-\alpha)c_2 \frac{m}{\operatorname{tr} S^{-1}} I_m \tag{3.9}$$

dominates the best linear estimator $c_2 S$ (where $c_2 = 1/(n+m+1)$) provided that $b_2/a_2 < \alpha < 1$ where

$$a_2 = [m(n-m+3) - 2(n+2)]c_2(n-m+1) + n, \tag{3.10}$$

$$b_2 = mc_2(n-m+3)(n-m+1) - n + 2m + 2. \tag{3.11}$$

PROOF. The proof is similar to the proof of Theorem 2.2. As n_2 tends to infinity, $n_2 F$ converges to a Wishart matrix with n_1 degrees of freedom and scale matrix Δ. This is same as the setting in Theorem 3.2 with n replace n_1 and Σ replace Δ. The results in Theorem 3.2 can be obtained from Theorem 3.1 by letting $n_2 \to \infty$ in (3.8). Note that

$$\lim_{n_2 \to \infty} c_4 a_1 = c_2 a_2 \quad \text{and} \quad \lim_{n_2 \to \infty} c_4 b_1 = c_2 b_2$$

where a_1, b_1 are defined in (3.5) and (3.6), and a_2, b_2 are defined in (3.10) and (3.11) respectively. Therefore the risk difference between $c_2 S$ and $\hat{\Sigma}_2(\alpha)$ is

$$G_3(\Sigma) \geq (1-\alpha)mc_2(a_2\alpha - b_2). \tag{3.12}$$

A sufficient condition for $G_3(\Sigma) > 0$ is $b_2/a_2 < \alpha < 1$. Note that $b_2/a_2 < 1$ iff $(n+2)(n-m+1) < (n-m-1)(n+m+1)$, which is equivalent to $n(m-3) > m^2 + 3$. Therefore

Table 2. PRIAL of $\hat{\Sigma}_1(\alpha)$ over $c_1 S$ using L_1 loss.

Σ	$n=5$	$n=10$	$n=25$
diag(1, 1, 1)	9.633	7.296	3.765
diag(4, 2, 1)	9.293	6.799	3.468
diag(25, 1, 1)	8.917	6.207	3.292

Table 3. PRIAL of $\hat{\Sigma}_2(\alpha)$ over $c_2 S$ using L_2 loss.

Σ	$n=20$	$n=25$	$n=50$
diag(1, 1, 1, 1)	0.110	0.450	0.651
diag(8, 4, 2, 1)	0.105	0.422	0.599
diag(25, 1, 1, 1)	0.106	0.430	0.632

the conditions $m > 3$ and $n > m + 3 + 12/(m-3)$ are needed to ensure $b_2/a_2 < 1$ as stated in (3.1). The conditions (3.2) and (3.3) are satisfied as $n_2 \to \infty$.

Remark 6. An 'optimal' value of α that maximizes the lower bound of $G_3(\Sigma)$ in (3.12) is

$$\begin{aligned}\alpha_3 &= (a_2 + b_2)/(2a_2) \\ &= \frac{[m(n-m+3) - n - 2](n-m+1) + (m+1)(n-m-1)}{[m(n-m+3) - 2n - 4](n-m+1) + n(n-m-1)}.\end{aligned}$$

Note also that $\alpha_3 < 1$ iff $b_2 < a_2$. Hence the conditions $n > m + 3 + 12/(m-3)$ and $m > 3$ in Theorem 3.2 ensure that $\alpha_3 < 1$.

4. Simulation study

For estimating the covariance matrix Σ in the Wishart distribution, a Monte Carlo simulation study was carried out to compare the risks of $\hat{\Sigma}_1(\alpha)$ and $c_1 S$ using L_1 loss with the 'optimal' value of α suggested in Remark 3. For $m = 3$ and $n = 5, 10, 25$, a sample of 1000 Wishart $W_3(n, \Sigma)$ matrices were generated for three different choices of Σ. Then these 1000 matrices were used to construct $c_1 S$ and $\hat{\Sigma}_1(\alpha)$ and the average losses (with respect to L_1) were obtained. Table 2 summarizes the percentage reduction in average loss (PRIAL) for $\hat{\Sigma}_1(\alpha)$ compared to $c_1 S$, i.e., it is the estimate of

$$\frac{E[L_1(\Sigma, c_1 S) - L_1(\Sigma, \hat{\Sigma}_1(\alpha))]}{E[L_1(\Sigma, c_1 S)]}.$$

Similarly we also compare the risks of $\hat{\Sigma}_2(\alpha)$ and $c_2 S$ using L_2 loss with the 'optimal' value of α suggested in remark 6. Due to the conditions stated in Theorem 3.2, we choose $m = 4$ and $n = 20, 25, 50$ in this simulation. Table 3 shows the PRIAL for $\hat{\Sigma}_2(\alpha)$ compared to $c_2 S$ using L_2 loss.

For estimating Δ in a multivariate F distribution, we generate 1000 random matrices of A's and B's from $W_m(n_1, \Delta)$ and $W_m(n_2, I)$ respectively for three different choices of

Table 4. PRIAL of $\hat{\Delta}_1(\alpha)$ over c_3F using L_1 loss.

Δ	$n_1 = n_2 = 5$	$n_1 = n_2 = 10$	$n_1 = n_2 = 25$
$\mathrm{diag}(1,1,1)$	7.508	9.662	6.364
$\mathrm{diag}(4,2,1)$	7.392	9.203	5.937
$\mathrm{diag}(25,1,1)$	7.253	8.678	5.620

Table 5. PRIAL of $\hat{\Delta}_2(\alpha)$ over c_4F using L_2 loss.

Δ	$n_1 = n_2 = 40$	$n_1 = n_2 = 45$	$n_1 = n_2 = 50$
$\mathrm{diag}(1,1,1,1,1,1,1,1)$	3.293	3.225	3.127
$\mathrm{diag}(128,64,32,16,8,4,2,1)$	3.101	3.015	2.882
$\mathrm{diag}(25,1,1,1,1,1,1,1)$	3.270	3.209	3.111

Δ. They are then transformed into $F = A^{1/2}B^{-1}A^{1/2}$. For simplicity, we restrict $n_1 = n_2$ in our simulation study. For the L_1 loss, we choose $m = 3$ and $n_1 = n_2 = 5, 10, 25$. Table 4 summarizes the PRIAL for $\hat{\Delta}_1(\alpha)$ compared to c_3F with the 'optimal' value of α suggested in Remark 1. For the L_2 loss, we choose $m = 8$ since $n_2 = n_1$ can assume reasonable values that satisfy the conditions in (3.3) and (3.4) as seen from Table 1. Therefore, we choose $m = 8$ and $n_1 = n_2 = 40, 45, 50$ in this study. The PRIAL for $\hat{\Delta}_2(\alpha)$ compared to c_4F with the "optimal" value of α suggested in Remark 4.

From Tables 2–5, all the PRIALs are positive and confirm the dominance results. Our proposed estimators represent a reasonably improvement over the corresponding best linear estimators except for estimating Σ in a Wishart distribution using the L_2 loss.

Acknowledgement

We would like to thank the referee whose comments improved the content and presentation of this paper.

Appendix

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 Δ, denoted by $F_m(n_1, n_2; \Delta)$. That is, F has the probability density function

$$\frac{\Gamma_m(n/2)}{\Gamma_m(n_1/2)\Gamma_m(n_2/2)} \quad |\Delta|^{-n_1/2}|F|^{(n_1-m-1)/2}|I + \Delta^{-1}F|^{-n/2}$$

where $n = n_1 + n_2$, $n_1 > m + 1$, $n_2 > m + 1$ and $\Gamma_m(\cdot)$ is the multivariate Gamma function. Let $V(F, \Delta)$ be a matrix whose elements are function of F and Δ and $V_{(r)} = rV + (1 - r)\,\mathrm{diag}(V)$. Note that $\mathrm{tr}(A_{(r)}B_{(1/r)}) = \mathrm{tr}(AB)$ for any $m \times m$ matrices A and B. We define

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

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 $\partial h(F)/\partial F$ is an $m \times m$ matrix whose (i,j)-th element is $\partial h(F)/\partial f_{ij}$. We simply write $V(F, \Delta)$ as V and $h(F)$ as h for brevity. Under fairly general regularity conditions, we have the F identity:

$$\text{(A.2)} \qquad E[h \operatorname{tr}(\Delta + F)^{-1}V] = \frac{2}{n}E[h \operatorname{tr}(DV)] + \frac{2}{n}E\left[\operatorname{tr}\left(\frac{\partial h}{\partial F}V_{(1/2)}\right)\right] + \frac{n_1 - m - 1}{n}E[h \operatorname{tr}(F^{-1}V)].$$

This F identity is an extension of the Wishart identity (Haff (1979)) to the multivariate F distribution. The regularity conditions are to ensure the function hV satisfies the conditions of the Stokes' theorem and are given in Konno (1988).

PROOF OF LEMMA 1. (i) Take $h(F) = 1/(\operatorname{tr} F^{-1})^2$ and $V = (\Delta + F)\Delta^{-2}$ in the F identity (A.2). Note that $\operatorname{tr} DV = [(m+1)/2] \operatorname{tr} \Delta^{-2}$ and $\partial h/\partial F = 2F_{(2)}^{-2}/(\operatorname{tr} F^{-1})^3$ (see Haff (1979)). Therefore,

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

Using $\operatorname{tr}[F^{-2}(\Delta + F)\Delta^{-2}] \le (\operatorname{tr} F^{-1})[\operatorname{tr}(F^{-1}(\Delta + F)\Delta^{-2})]$ and simplify,

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

It is same as equation (A.1) in Konno (1991). Using

$$\text{(A.3)} \qquad \operatorname{tr}(F^{-1}\Delta^{-1}) \le (\operatorname{tr} F^{-1})(\operatorname{tr} \Delta^{-1}),$$

the result (i) follows immediately.

(ii) Take $h(F) = 1/(\operatorname{tr} F^{-1})$ and $V = (\Delta + F)\Delta^{-2}F$ in (A.2). $\partial h/\partial F = F_{(2)}^{-2}/(\operatorname{tr} F^{-1})^2$ (see Haff (1979)) and using Lemma (2.2) in Konno (1988), it can be shown that $\operatorname{tr} DV = [(m+1)/2](\operatorname{tr} \Delta^{-1}) + (m+1)\operatorname{tr}(\Delta^{-2}F)$.

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

Applying (A.3) in the second term and (i) in the last term of the right hand side of (A.4), the result (ii) follows after simplification.

(iii) Take $h(F) = 1/(\operatorname{tr} F^{-1})$ and $V = (\Delta + F)\Delta^{-1}$ in (A.2). Note that $\operatorname{tr} DV = [(m+1)/2](\operatorname{tr} \Delta^{-1})$ and $\partial h/\partial F = F_{(2)}^{-2}/(\operatorname{tr} F^{-1})^2$ (see Haff (1979)).

$$E\left[\frac{\operatorname{tr} \Delta^{-1}}{\operatorname{tr} F^{-1}}\right] = \frac{n_1 - m - 1}{n_2} + \frac{2}{n_2} E\left[\frac{\operatorname{tr} F^{-2}}{(\operatorname{tr} F^{-1})^2}\right] + \frac{2}{n_2} E\left[\frac{\operatorname{tr}(F^{-1}\Delta^{-1})}{(\operatorname{tr} F^{-1})^2}\right]. \tag{A.5}$$

Note that the last two terms of (A.5) are nonnegative, the lower bound in (iii) follows. For the upper bound in (iii), we apply (A.3) to the last term of (A.5) and $\operatorname{tr} F^{-2}/(\operatorname{tr} F^{-1})^2 \leq 1$ to the second term of (A.5), which completes the proof.

References

Dey, D. K. (1988). Simultaneous estimation of eigenvalues, *Ann. Inst. Statist. Math.*, **40**, 137–147.

Dey, D. K. (1989). On estimation of the scale matrix of the multivariate F distribution, *Comm. Statist. Theory Methods*, **18**, 1373–1383.

Dey, D. K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss, *Ann. Statist.*, **13**, 1581–1591.

Friedman, J. H. (1989). Regularized discriminant analysis, *J. Amer. Statist. Assoc.*, **84**, 165–175.

Gupta, A. K. and Krishnamoorthy, K. (1990). Improved estimators of eigenvalues of $\Sigma_1\Sigma_2^{-1}$, *Statist. Decisions*, **8**, 247–263.

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

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

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

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

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. and Chan, W. Y. (1998). Estimation of the scale matrix and its eigenvalues in the Wishart and the Multivariate F distributions, *Ann. Inst. Statist. Math.*, **50**, 523–530.

Leung, P. L., and Muirhead, R. J. (1988). On estimating characteristic roots in a two–sample problem, *Statistical Decision Theory and Related Topics IV*, Vol. 1 (eds. S. S. Gupta and J. O. Berger), 397–402, Springer, New York.

Lin, S. P. and Perlman, M. D. (1985). A Monte Carlo comparison of four estimators for a covariance matrix, *Multivariate Analysis VI* (ed. P. R. Krishnaiah), 411–429, North Holland, Amsterdam.

Merritt, F. S. (1962). *Mathematics Manual*, McGraw–Hill. New York.

Muirhead, R. J. (1987). Developments in eigenvalue estimation, *Advances in Multivariate Statistical Analysis*, 277–288, Reidel, Dordrecht.

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

Pal, N. (1993). Estimating the normal dispersion matrix and the precision matrix from a decision–theoretic point of view: A review, *Statist. Papers*, **34**, 1–26.

Sheena, Y. and Takemura, A. (1992). Inadmissibility of non-order-preserving othogonally invariant estimators of the covariance matrix in the case of Stein's loss, *J. Multivariate Anal.*, **41**, 117–131.

Stein, C. (1975). Estimation of a covariance matrix, *Rietz Lecture*, *39th Annual meeting IMS*, *Atlanta, Georgia*.