JOURNAL OF MULTIVARIATE ANALYSIS **41**, 1–13 (1992)

Estimation of the Eigenvalues of $\Sigma_1 \Sigma_2^{-1}$

M. BILODEAU

Université de Montréal, Montréal, Canada

AND

M. S. SRIVASTAVA

University of Toronto, Toronto, Canada

Communicated by the Editors

In the normal two-sample problem, an invariant test for the hypothesis of the equality of the population covariance matrices, $H : \Sigma_1 = \Sigma_2$ vs $A : \Sigma_1 \neq \Sigma_2$, has a power function which depends only on the eigenvalues of $\Sigma_1 \Sigma_2^{-1}$. An orthogonally invariant minimax estimator of these eigenvalues is proposed which has very desirable properties. Namely, the estimated eigenvalues are always positive and they follow the same ordering as the eigenvalues of $S_1 S_2^{-1}$ calculated from the usual sample covariance matrices. Moreover, it has an explicit expression that can be easily calculated and yields substantial risk reductions.

1. INTRODUCTION

Let S_1 and S_2 be two independent Wishart matrices distributed as $W_p(\Sigma_i, n_i)$, $n_i > p+1$, $i = 1, 2$. The problem of testing the equality of the covariance matrices of two normal populations is left invariant by the group of transformations

$$G = \{(B, c, d);\ B \in \mathrm{Gl}(p),\ c \in \mathbb{R}^p,\ d \in \mathbb{R}^p\}$$

acting on the sample space as

$$(\bar{x}_1, \bar{x}_2, S_1, S_2) \to (B\bar{x}_1 + c,\ B\bar{x}_2 + d,\ BS_1B',\ BS_2B').$$

Received February 13, 1991; revised September 2, 1991.

AMS 1980 subject classification numbers: primary 62F10; secondary 62C99.

Key words and phrases: decision theory, loss function, minimax estimator, multivariate F-distribution, orthogonally invariant estimator, unbiased risk estimator.

0047-259X/92 \$3.00

Moreover, the power function of an invariant test depends on the parameter only through the maximal invariant parameters $(\delta_1, \ldots, \delta_p)$, where $\delta_1 \geqslant \cdots \geqslant \delta_p$ are the eigenvalues of $\Sigma_1 \Sigma_2^{-1}$. Our concern is the estimation of those eigenvalues using a decision-theoretic approach. The distribution of the eigenvalues of $S_1 S_2^{-1}$ is very involved and does not lend itself to the derivation of improved estimators. Instead Muirhead and Verathaworn [9] proposed the following approach, which is the one taken here.

Let H be a non-singular matrix such that $H\Sigma_2 H' = I$ and $\Delta = H\Sigma_1 H'$. Define $A = HS_1 H' \sim W_p(\Delta, n_1)$ and $B = HS_2 H' \sim W_p(I, n_2)$. The eigenvalues of Δ are the same as those of $\Sigma_1 \Sigma_2^{-1}$ and the eigenvalues of $S_1 S_2^{-1}$ are the same as those of the nonobservable matrix

$$F = A^{1/2} B^{-1} A^{1/2}$$

distributed as a multivariate $F_p(n_1, n_2; \Delta)$ distribution. The approach consists of estimating the eigenvalues δ_i by the eigenvalues of an orthogonally invariant estimator $\hat{\Delta}(F) = R\Phi(L)R'$, where $F = RLR'$, $RR' = I$, $L = \operatorname{diag}(l_1, \ldots, l_p)$ and $\Phi(L) = \operatorname{diag}(\phi_1, \ldots, \phi_p)$. It should be noted that the eigenvalues of $\hat{\Delta}(F)$ depends on F only through $l_1, \ldots, l_p$ and hence they are proper estimates.

Let F be a $p \times p$ positive definite random matrix with probability density function

$$\frac{\Gamma_p(n/2)}{\Gamma_p(n_1/2)\,\Gamma_p(n_2/2)} |\Delta|^{-n_1/2} |F|^{(n_1 - p - 1)/2} |I + \Delta^{-1} F|^{-n/2}, \qquad F > 0, \tag{1.1}$$

where $n = n_1 + n_2$. Denote by $F \sim F_p(n_1, n_2; \Delta)$ the distribution with density (1.1). The problem of estimating the eigenvalues of the positive definite matrix Δ has been considered by some authors. Muirhead and Verathaworn [9] considered the problem of estimating Δ by $\hat{\Delta}(F)$, using the invariant loss function

$$\mathscr{L}_S(\Delta, \hat{\Delta}) = \operatorname{tr}(\Delta^{-1}\hat{\Delta}) - \log |\Delta^{-1}\hat{\Delta}| - p. \tag{1.2}$$

This loss function was first proposed for the estimation of the eigenvalues of the scale matrix of a Wishart distribution. However, for the estimation of the scale matrix Δ of a multivariate F-distribution (1.1) the derivation of improved estimators is not as easy.

Muirhead and Verathaworn [9] obtained an approximate expression for the variational form of the Bayes estimate based on an approximation to the unbiased estimate of the risk in conjunction with the Euler–Lagrange system of differential equations [3, 4]. It has not been established that the estimators proposed have a frequentist risk uniformly smaller than the best multiple of F which is just the unbiased estimate

$$\hat{\Delta}_U = \frac{n_2 - p - 1}{n_1} F. \tag{1.3}$$

In their Table 1 a simulation study also showed that $\hat{\Delta}_U$ suffers the same defects as in the one sample problem where the best multiple of the covariance matrix is used to estimate the population covariance matrix. Explicitly, the latent roots tend to be much more dispersed than those of Δ. The smallest eigenvalues are biased too low while the largest eigenvalues are biased in the opposite direction. These biases become critical especially for small values of n_1 and n_2. More recently, Konno [6] considered orthogonally invariant estimators of the form

$$\hat{\Delta}_K = a_1(F + ut(u)I), \tag{1.4}$$

where $a_1 = (n_2 - p - 1)/n_1$ and $t(u)$ is an absolutely continuous and nonnegative function of $u = 1/\text{tr}(F^{-1})$. He showed directly, without the unbiased risk estimate, that for $p \geqslant 2$ if the function $t(u)$ is nonincreasing and is bounded as $0 \leqslant t(u) \leqslant 2(p-1)(n-p-1)/n_1(n_2-2)$ then $\hat{\Delta}_K$ has uniformly smaller risk than $\hat{\Delta}_U$. No simulation study was done to assess the risk reduction, but since all eigenvalues of F are modified in the same direction, no substantial risk reduction should be expected.

In this paper, the estimation of Δ is done with respect to the invariant loss function

$$\mathscr{L}_F(\Delta, \hat{\Delta}) = \text{tr}[(\Delta + F)^{-1}(\hat{\Delta} + F)] - \log|(\Delta + F)^{-1}(\hat{\Delta} + F)| - p. \tag{1.5}$$

Section 2 deals with the unbiased risk estimate of a nearly arbitrary estimator $\hat{\Delta}(F)$ and gives the best multiple estimator. This unbiased risk estimator is specialized to orthogonally invariant estimators. In Section 3 the exact form of the orthogonally invariant Bayes rule as the solution to the Euler–Lagrange system of differential equations is obtained. An improved estimator which modifies all eigenvalues of the best multiple estimator in the same direction similar to those of Haff for the eigenvalues of the mean of a Wishart distribution is given in Section 4. Since the best multiple estimator is not minimax it cannot be concluded that the improved estimators are minimax. In Section 5, the minimax estimator which is the best equivariant estimator with respect to the group G_T^+ of lower triangular matrices with positive diagonal elements is constructed. This estimator is then modified "à la Eaton" [1] to obtain improved minimax estimators that are orthogonally invariant. Finally, some simulations reported in Section 6 show that the orthogonally invariant minimax estimator offers substantial risk improvements for various degrees of freedom and matrices Δ.

2. Unbiased Estimate of the Risk

Following Haff [4] we would like to obtain an unbiased estimate of the risk of a nearly arbitrary estimator $\hat{\Delta}(F)$. This can be done using the F-identity of Muirhead and Verathaworn [9].

Let D be the $p \times p$ matrix of partial derivative operators whose (i, j) element is $d_{ij} = \frac{1}{2}(1 + \delta_{ij})\ \partial/\partial f_{ij}$, where δ_{ij} is the Kronecker delta and $F = (f_{ij})$. For a matrix function $Q \equiv Q(F) : p \times p$ define the matrix $DQ = (\sum_{k=1}^{p} d_{ik} Q_{kj})$.

The F-Identity. Let $V = (v_{ij}(F)) : p \times p$ satisfy the conditions stated in Verathaworn [14]. Then

$$E\{\operatorname{tr}(\Delta + F)^{-1} V\} = E\left\{\frac{2}{n} \operatorname{tr}(DV) + \frac{n_1 - p - 1}{n} \operatorname{tr}(F^{-1}V)\right\}. \tag{2.1}$$

The risk of an estimator $\hat{\Delta}(F)$ is

$$\begin{aligned} R(\Delta, \hat{\Delta}) &= E\{\operatorname{tr}[(\Delta + F)^{-1} (\hat{\Delta} + F)] - \log |(\Delta + F)^{-1} (\hat{\Delta} + F)| - p\} \\ &= E\{\operatorname{tr}[(\Delta + F)^{-1} \hat{\Delta}] - \log |\hat{\Delta} + F| + \operatorname{tr}[(\Delta + F)^{-1} F] \\ &\quad - \log |(\Delta + F)^{-1}| - p\}. \end{aligned}$$

If the constant terms are omitted, equivalently the risk

$$R^*(\Delta, \hat{\Delta}) = E\{\operatorname{tr}[(\Delta + F)^{-1} \hat{\Delta}] - \log |\hat{\Delta} + F|\} \tag{2.2}$$

can be considered. An unbiased estimate of R^* is readily obtained from the F-identity,

$$\hat{R}^*(\Delta, \hat{\Delta}) = \frac{2}{n} \operatorname{tr}(D\hat{\Delta}) + \frac{n_1 - p - 1}{n} \operatorname{tr}(F^{-1}\hat{\Delta}) - \log |\hat{\Delta} + F|. \tag{2.3}$$

The best multiple estimator aF with constant risk is now easily obtained. The unbiased estimator of $R^*(\Delta, aF)$ is

$$\begin{aligned} \hat{R}^*(\Delta, aF) &= 2a \operatorname{tr}(DF)/n + (n_1 - p - 1)\, ap/n - p \log(a + 1) - \log |F| \\ &= ap(p + 1)/n + (n_1 - p - 1)\, ap/n - p \log(a + 1) - \log |F| \\ &= apn_1/n - p \log(a + 1) - \log |F| \end{aligned}$$

which is minimized at $a = n_2/n_1$. Therefore, one possibility is to seek improvements on the risk of the best multiple estimator

$$\hat{\Delta}_0 = \frac{n_2}{n_1} F.$$

From now on, we restrict our attention to orthogonally invariant estimators with the same eigenvectors as $\hat{\Delta}_0$ but modified (and observable) eigenvalues.

Let $F = RLR'$, where R is orthogonal and $L = \mathrm{diag}(l)$, $l = (l_1, ..., l_p)'$, $l_1 \geqslant \cdots \geqslant l_p$. Orthogonally invariant estimators are of the form

$$\hat{\Delta}_I = R\Phi(L)R',$$

where $\Phi(L) = \mathrm{diag}(\phi)$, and $\phi = (\phi_1, ..., \phi_p)'$.

LEMMA 1 [4, 12]. *Let $\phi_i(L)$, $i = 1, ..., p$, be differentiable on $l_1 > \cdots > l_p$. Then*

$$D(R\Phi(L)R') = R\Phi^{(1)}(L)R',$$

in which

$$\Phi^{(1)}(L) = \mathrm{diag}(\phi_1^{(1)}(L), ..., \phi_p^{(1)}(L)),$$

$$\phi_i^{(1)}(L) = \frac{1}{2} \sum_{\substack{j=1 \\ j \neq i}}^{p} \frac{\phi_i(L) - \phi_j(L)}{l_i - l_j} + \partial\phi_i(L)/\partial l_i.$$

From Lemma 1 the unbiased estimate of the risk of $R\Phi(L)R'$ follows.

THEOREM 1. *The unbiased estimator of the risk $R^*(\Delta, R\Phi R')$ is given by*

$$\begin{aligned}\hat{R}^*(\Delta, R\Phi R') &= \frac{1}{n} \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \frac{\phi_i - \phi_j}{l_i - l_j} + \frac{2}{n} \sum_{i=1}^{p} \partial\phi_i/\partial l_i \\ &\quad + \frac{n_1 - p - 1}{n} \sum_{i=1}^{p} \phi_i/l_i - \sum_{i=1}^{p} \log(\phi_i + l_i) \\ &= \hat{R}^*(l, \phi, d),\end{aligned}$$

where $d = \nabla \cdot \phi = \sum_{i=1}^{p} \partial\phi_i/\partial l_i$ is the divergence of ϕ.

3. Formal Bayes Estimators

Define $\lambda = (\lambda_1, ..., \lambda_p)'$ the eigenvalues of Δ. Let $\pi(\Delta) = \pi^*(\lambda)$ be an orthogonally invariant prior distribution (i.e., $\pi(\Gamma\Delta\Gamma') = \pi(\Delta)$ for any orthogonal matrix Γ). Denote by $f(\mathbf{l} \mid \lambda)$ the conditional density of $\mathbf{l} = (l_1, ..., l_p)'$, given $\lambda = (\lambda_1, ..., \lambda_p)'$. Finally, the marginal density of $\mathbf{l}$ is denoted by

$$g_\pi(\mathbf{l}) = \int f(\mathbf{l} \mid \lambda)\, d\pi^*(\lambda). \tag{3.1}$$

Following Haff [4] the Bayes risk of the estimator $R\Phi R'$ is given by

$$r(\phi, \pi) = \int \hat{R}^*(\mathbf{l}, \phi, d)\, g_\pi(\mathbf{l})\, d\mathbf{l}. \tag{3.2}$$

Since the loss function is convex, then the formal Bayes rule is unique and is obtained by minimizing the functional $r(\phi, \pi)$. The function ϕ minimizing that functional must satisfy the Euler–Lagrange differential equations

$$\partial \hat{R}^*/\partial \phi_i = \frac{\partial}{\partial l_i} (\partial \hat{R}^*/\partial d) + (\partial \hat{R}^*/\partial d)(\partial \log g_\pi/\partial l_i), \qquad i = 1, ..., p. \tag{3.3}$$

Now, since $\partial \hat{R}^*/\partial d = 2/n$ the first term on the right side of (3.3) vanishes and the Euler–Lagrange equations reduce to

$$\frac{2}{n} \sum_{\substack{j=1 \\ j \neq i}}^{p} \frac{1}{l_i - l_j} + \frac{(n_1 - p - 1)}{n} \frac{1}{l_i} - \frac{1}{\phi_i + l_i} = \frac{2}{n}\, \partial \log g_\pi/\partial l_i, \qquad i = 1, ..., p. \tag{3.4}$$

It is readily checked that the solution of this system is given by

$$\phi_i = n l_i \left[(n_1 - p - 1) + 2 l_i \sum_{\substack{j=1 \\ j \neq i}}^{p} \frac{1}{l_i - l_j} - 2 l_i \partial \log g_\pi(\mathbf{l})/\partial l_i \right]^{-1} - l_i. \tag{3.5}$$

This is the exact Bayes estimate of the vector of eigenvalues of Δ. They depend on the prior distribution $\pi^*(\lambda)$ through the marginal density $g_\pi(\mathbf{l})$.

The variational form of the Bayes estimator (3.5) was derived using an unbiased estimator of the risk $\hat{R}^*$ of an orthogonally invariant estimator. Of course, the posterior expected loss can be minimized directly to obtain another form of the Bayes estimator with respect to the prior $\pi(\Delta)$ (not necessarily orthogonally invariant). The Bayes rule with respect to the loss function $\mathscr{L}_F$ is $\hat{\Delta}(F) = \{E^{\pi(\Delta \mid F)}[(\Delta + F)^{-1}]\}^{-1} - F$.

4. Estimators Improving upon the Best Multiple

Theorem 2. *Let $p \geqslant 2$. The estimator $R\Phi R'$, where*

$$\begin{aligned} \phi_i &= \frac{n_2}{n_1}(l_i + ut(u)), \\ u &= 1/\operatorname{tr}(F^{-1}) \end{aligned} \tag{4.1}$$

has uniformly smaller risk than the best multiple estimator if $t(u)$ is a nonincreasing function such that $0 \leqslant t(u) \leqslant 2(p-1)(1/n_1 + 1/n_2)$.

Proof. Since $t'(u) \leqslant 0$, the divergence of ϕ can be bounded above as

$$\begin{aligned} \sum_{j=1}^{p} \partial\phi_j/\partial l_j &= \frac{n_2}{n_1}\left[p + \left(\sum_{j=1}^{p} l_j^{-2}\right) u^3 t'(u) + \left(\sum_{j=1}^{p} l_j^{-2}\right) u^2 t(u)\right] \\ &\leqslant \frac{n_2}{n_1}(p + t(u)). \end{aligned}$$

Therefore an upper bound for the risk difference is

$$\begin{aligned} &\hat{R}^*(\Delta, R\Phi R') - \hat{R}^*(\Delta, \hat{\Delta}_0) \\ &\quad \leqslant \frac{1}{n}\frac{n_2}{n_1} p(p-1) + \frac{2}{n}\frac{n_2}{n_1}(p + t(u)) + \frac{n_1 - p - 1}{n}\frac{n_2}{n_1}(p + t(u)) \\ &\qquad - \sum_{j=1}^{p}\left[\log\left(\frac{n_2}{n_1} + 1\right) l_j + \frac{n_2}{n_1} ut(u)\right] - \frac{n_2}{n} p + \sum_{j=1}^{p} \log\left(\frac{n_2}{n_1} + 1\right) l_j \\ &\quad = \frac{n_2}{n_1}\frac{t(u)}{n}(n_1 - p + 1) - \sum_{j=1}^{p} \log\left[1 + \frac{n_2}{n} t(u) l_j^{-1} u\right] \\ &\quad \leqslant \frac{n_2}{n} t(u)\left[\frac{-(p-1)}{n_1} + \frac{t(u)}{2}\frac{n_2}{n}\right] \leqslant 0 \end{aligned}$$

if $0 \leqslant t(u) \leqslant 2(p-1)(1/n_1 + 1/n_2)$.

The problem with the best multiple estimator is the overdispersion of the eigenvalues. Small eigenvalues are biased too low while large eigenvalues are biased in the opposite direction. The estimator proposed in (4.1) will not correct this problem, since all eigenvalues are moved in the same direction. Substantial risk reduction should not be expected and better estimators that really correct the problem described should be sought.

5. Minimax Estimators

Let $F = TT'$ be the Bartlett decomposition of F and $D = \mathrm{diag}(d_1, ..., d_p)$. The estimators equivariant with respect to the group G_T^+ of triangular matrices with positive diagonal elements are of the form $\hat{\Delta} = TDT'$. This group being solvable the best equivariant estimator is minimax.

Lemma 2. *Let $G = T'(I+F)^{-1} T$, where $F = TT'$ is the Bartlett decomposition of $F \sim F_p(n_1, n_2; I)$. Then the expectation of the diagonal elements of $G = (g_{ij})$ are given by $E(g_{ii}) = (n_1 + p - 2i + 1)/n$.*

Proof. The Jacobian from F to T is $J(F \to T) = 2^p \prod_{i=1}^p t_{ii}^{(p-i+1)}$. Thus it follows that the density of T is

$$p(T) = c2^p \prod_{i=1}^{p} t_{ii}^{n_1 - i} / |I + TT'|^{n/2},$$

where c is the normalizing constant. Now, partition T as

$$T = \begin{pmatrix} t_{11} & \mathbf{0}' \\ \mathbf{t}_{21} & T_{22} \end{pmatrix},$$

where t_{11} is a scalar.

It can be shown that

$$|I + TT'| = (1 + t_{11}^2)\, |I + T_{22} T_{22}'|\, |I + \mathbf{z}\mathbf{z}'|,$$

where

$$\mathbf{z} = \frac{(I + T_{22} T_{22}')^{-1/2}}{(1 + t_{11}^2)^{1/2}} \mathbf{t}_{21}.$$

Furthermore, since $J(\mathbf{t}_{21} \to \mathbf{z}) = (1 + t_{11}^2)^{(p-1)/2} |I + T_{22} T_{22}'|^{1/2}$, the joint density of t_{11}, T_{22}, and $\mathbf{z}$ is given by

$$\frac{c2^p \prod_{i=1}^p t_{ii}^{n_1 - i}}{(1 + t_{11}^2)^{(n-p+1)/2}} |I + T_{22} T_{22}'|^{-(n-1)/2} (1 + \mathbf{z}'\mathbf{z})^{-n/2}.$$

Hence, t_{11}, T_{22}, and $\mathbf{z}$ are independent with

$$\begin{aligned} t_{11}^2 &\sim F(n_1, n_2 - p + 1), \\ W = T_{22} T_{22}' &\sim F_{p-1}(n_1 - 1, n_2; I), \qquad (5.1) \\ w = \mathbf{z}'\mathbf{z} &\sim F(p-1, n-p+1). \end{aligned}$$

The distributions of t_{11}^2 and w are canonical F-distributions [2] and the distribution of w follows from general results on spherical distributions [11, Lemma 3.2.3].

Now, since $G = I - (I + T'T)^{-1}$ it follows that $g_{11} = 1 - (1 + t_{11}^2)^{-1} (1 + w)^{-1}$ whose expectation is $E(g_{11}) = (n_1 + p - 1)/n$. The matrix G can also be partitioned as

$$\begin{pmatrix} g_{11} & \mathbf{g}_{21}' \\ \mathbf{g}_{21} & G_{22} \end{pmatrix},$$

where

$$\begin{aligned} G_{22} &= T_{22}'\{(I + T_{22}T_{22}')^{1/2} (I + \mathbf{z}\mathbf{z}')(I + T_{22}T_{22}')^{1/2}\}^{-1} T_{22} \\ &= T_{22}'(I + T_{22}T_{22}')^{-1} T_{22} \\ &\quad - T_{22}'(I + T_{22}T_{22}')^{-1/2} \frac{\mathbf{z}\mathbf{z}'}{(1 + \mathbf{z}'\mathbf{z})} (I + T_{22}T_{22}')^{-1/2} T_{22}. \end{aligned}$$

Because the distribution of $\mathbf{z}$ is spherical then necessarily $E(\mathbf{z}\mathbf{z}'/(1 + \mathbf{z}'\mathbf{z})) = \alpha I_{p-1}$ for a certain constant α. Therefore,

$$E(G_{22}) = (1 - \alpha)\, E\{T_{22}'(I + T_{22}T_{22}')^{-1} T_{22}\}.$$

The constant α is easily determined from $E(w/(1 + w)) = \alpha(p - 1) = (p - 1)/n$ from which $\alpha = 1/n$. The expectation of G_{22} has the same form as the expectation of G but in dimension $(p - 1)$. The result follows by induction.

THEOREM 3. *Let $F = TT'$. The best equivariant estimator $\hat{\Delta} = TDT'$ of Δ is given by $D = \operatorname{diag}(d_1, ..., d_p)$, $d_1 < \cdots < d_p$, where*

$$d_i = \frac{n_2 - p + 2i - 1}{n_1 + p - 2i + 1}, \qquad i = 1, ..., p. \tag{5.2}$$

The minimax risk is

$$R^*(I, TDT') = pn_2/n - \sum_{i=1}^{p} \{\log(d_i + 1) + E[\log(F_{n_1 - i + 1, n_2 - p + i})]\}.$$

Proof. For $\Delta = I$,

$$\begin{aligned} R^*(I, \hat{\Delta}) &= E\left\{\operatorname{tr}[(I + F)^{-1} TDT'] - \sum_{i=1}^{p} \log(d_i + 1) - \log |F|\right\} \\ &= E\left\{\operatorname{tr}(GD) - \sum_{i=1}^{p} \log(d_i + 1) - \log |F|\right\}, \end{aligned}$$

where $G = T'(I + F)^{-1} T$. A differentiation with respect to d_i gives

$E(g_{ii})-(d_i+1)^{-1}=0$. It follows that the optimal choice is $d_i = E^{-1}(g_{ii})-1$. This reduces to (5.2) using the preceding lemma. For the minimax risk, it can be shown by induction using (5.1) that t_{ii}^2 follows a canonical $F(n_1-i+1,\ n_2-p+i)$ distribution, $i=1, ..., p$. This completes the proof.

The estimator in Theorem 3 has the unappealing feature that it depends on the coordinate system. Eaton [1] provided a way of obtaining minimax estimators that are orthogonally invariant in the context of estimating the mean of a Wishart distribution. His approach is as follows. Let $S \sim W_p(\Sigma, n)$ and $S=TT'$ be the Bartlett decomposition. Let $\hat{\Sigma}_M(S)=TDT'$ be the best equivariant estimator with respect to G_T^+. The matrix $D=\text{diag}(\mathbf{d})$ depends on the strict convex loss function considered. For any $H \in \text{Gl}(p)$ define

$$(H^*\hat{\Sigma}_M)(S)=H^{-1}\hat{\Sigma}_M(HSH')\,H^{-1\prime}.$$

The estimator proposed by Eaton is

$$\hat{\Sigma}_E(S)=E[(H^*\hat{\Sigma}_M)(S)\mid S] \tag{5.4}$$

and he suggested using a random matrix H that has the uniform distribution on the orthogonal group $\mathcal{O}(p)$ and is independent of S. The estimator (5.4) is then orthogonally invariant and minimax. It has the form $\Sigma_E(S)=R\Phi(L)\,R'$, where $S=RLR'$, and $\Phi(L)=\text{diag}(\phi(L))$. Takemura [13] gave a series expansion for $p=3$ and provided the factorization $\phi=LW(L)\mathbf{d}$, where $W(L)$ is doubly stochastic. The explicit calculations of $W(L)$ for $p>2$ remains an intractable problem. Perron [10] gave an approximation to the matrix $W(L)$, say $\tilde{W}(L)$ described below, and showed that the estimator $\hat{\Sigma}_P(S)=R\,\text{diag}(\tilde{\phi})R'$, where $\tilde{\phi}=L\tilde{W}\mathbf{d}$, is minimax and, moreover, $\tilde{\phi}$ and $\mathbf{l}$ follow the same order relation with the elements of $\tilde{\phi}$ being positive. It will be shown that the matrix $\tilde{W}$ serves an identical role for the estimation problem of the latent roots of $\Sigma_1\Sigma_2^{-1}$. This should not be surprising, since the exact expression for $W(L)$ is based on the conditional expectation given S in (5.4). At this point the Wishart distribution plays no role.

The matrix $\tilde{W}$ is defined in terms of

$$L_i=\text{diag}(l_1, ..., l_{i-1}, 0, l_{i+1}, ..., l_p), \qquad i=1, ..., p,$$

and

$$\text{tr}_k(L)=\begin{cases} 1, & \text{if } k=0 \\ \displaystyle\sum_{1\leqslant i_1<\cdots<i_k\leqslant p}\ \prod_{j=1}^{k} l_{i_j} & \text{if } k=1, ..., p, \\ 0 & \text{otherwise.}\end{cases}$$

The matrix $\tilde{W}$ is given componentwise by

$$\tilde{w}_{ik} = \mathrm{tr}_{k-1}^{-1}(L)\,\mathrm{tr}_{k-1}(L_i) - \mathrm{tr}_k^{-1}(L)\,\mathrm{tr}_k(L_i). \tag{5.5}$$

LEMMA 3. [10]. (i) *$\tilde{W}$ is doubly stochastic.*

(ii) $l_i(\partial/\partial l_i)[\mathrm{tr}_k(L_i)/\mathrm{tr}_k(L)] = -(\mathrm{tr}_k(L_i)/\mathrm{tr}_k(L))(1 - \mathrm{tr}_k(L_i)/\mathrm{tr}_k(L)) \leqslant 0.$

(iii) $\sum_{i>j} (l_i \tilde{w}_{ik}(L) - l_j \tilde{w}_{jk}(L))/(l_i - l_j) = (p-k).$

(iv) *Let* $\psi = \tilde{W}\mathbf{d}$, *if* $d_1 < \cdots < d_p$ *then* $d_1 < \psi_1 < \cdots < \psi_p < d_p$.

(v) *If* $l_1 > \cdots > l_p$ *then* $\tilde{\phi}_1 > \cdots > \tilde{\phi}_p$.

THEOREM 4. *The orthogonally invariant estimator* $\hat{\Delta} = R\ \mathrm{diag}(\tilde{\phi})R'$, *where* $\tilde{\phi} = L\tilde{W}\mathbf{d}$, $\mathbf{d}$ *as in* (5.2), *is minimax. Moreover,* $\tilde{\phi}$ *and* $\boldsymbol{l}$ *have the same order relation and the components of* $\tilde{\phi}$ *are all positive.*

Proof. Let $\psi = \tilde{W}\mathbf{d}$. Then, from Theorems 1 and 3,

$$\hat{R}^*(\Delta, R\ \mathrm{diag}(\tilde{\phi})R') - \hat{R}^*(I, TDT')$$

$$= \frac{1}{n}(n_1 - p + 1)\sum_{i=1}^{p} \psi_i + \frac{1}{n}\sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p} \frac{l_i\psi_i - l_j\psi_j}{l_i - l_j} - \frac{p}{n} n_2$$

$$+ \frac{2}{n}\sum_{i=1}^{p} l_i\,\partial\psi_i/\partial l_i - \sum_{i=1}^{p} [\log(\psi_i + 1) - \log(d_i + 1)]$$

$$= \frac{1}{n}(n_1 - p + 1)\sum_{k=1}^{p} d_k + \frac{2}{n}\sum_{k=1}^{p}(p-k)\,d_k - \frac{p}{n} n_2 \tag{5.6}$$

$$-\frac{2}{n}\sum_{i=1}^{p}\sum_{k=1}^{p-1} \frac{\mathrm{tr}_k(L_i)}{\mathrm{tr}_k(L)}\left(1 - \frac{\mathrm{tr}_k(L_i)}{\mathrm{tr}_k(L)}\right)(d_{k+1} - d_k) \tag{5.7}$$

$$-\sum_{i=1}^{p}\left[\log\left(\sum_{k=1}^{p} \tilde{w}_{ik}(d_k + 1)\right) - \sum_{k=1}^{p} \tilde{w}_{ik}\log(d_k + 1)\right] \tag{5.8}$$

$$= A + B + C,$$

where A, B, C are given respectively by (5.6), (5.7), (5.8). From the expression of d_k in (5.2) it follows that $A = 0$. The Lemma 3(ii) implies that $B \leqslant 0$ and finally the concavity of the logarithmic function gives $C \leqslant 0$. This completes the proof.

6. SIMULATION RESULTS

Table I gives the estimated risk of four estimators with respect to the loss function $\mathscr{L}_F$. Each simulated risk is an average of 500 realisations of the

TABLE I

Risks Comparisons

$n_1 = n_2$	10	25	50	100
Minimax risk	0.5803 (0.0150)	0.2059 (0.0050)	0.0979 (0.0021)	0.0497 (0.0010)
$\hat{\Delta}_0$ (best multiple)	0.6414 (0.0158)	0.2240 (0.0055)	0.1063 (0.0022)	0.0504 (0.0011)
$\Delta = \mathrm{diag}(1, 1, 1, 1)$				
$\hat{\Delta}_K$	0.5597 (0.0148) [12.7%]	0.1969 (0.0045) [12.1%]	0.0987 (0.0021) [7.1%]	0.0505 (0.0010) [−0.2%]
$\hat{\Delta}$	0.4169 (0.0136) [35.0%]	0.1724 (0.0042) [23.0%]	0.0922 (0.0020) [13.3%]	0.0489 (0.0011) [3.1%]
$\Delta = \mathrm{diag}(25, 1, 1, 1)$				
$\hat{\Delta}_K$	0.5368 (0.0130) [16.3%]	0.1974 (0.0044) [11.9%]	0.0976 (0.0020) [8.2%]	0.0500 (0.0010) [0.8%]
$\hat{\Delta}$	0.4334 (0.0120) [32.4%]	0.1823 (0.0042) [18.6%]	0.0940 (0.0020) [11.6%]	0.0490 (0.0010) [2.8%]
$\Delta = \mathrm{diag}(10, 10, 1, 1)$				
$\hat{\Delta}_K$	0.5547 (0.0139) [13.5%]	0.2027 (0.0046) [9.5%]	0.1039 (0.0022) [2.3%]	0.0533 (0.0011) [−5.7%]
$\hat{\Delta}$	0.4499 (0.0129) [29.9%]	0.1854 (0.0044) [17.2%]	0.0996 (0.0021) [6.3%]	0.0523 (0.0011) [−3.6%]
$\Delta = \mathrm{diag}(8, 4, 2, 1)$				
$\hat{\Delta}_K$	0.5551 (0.0131) [13.4%]	0.2097 (0.0047) [6.4%]	0.0991 (0.0022) [6.8%]	0.0501 (0.0010) [0.6%]
$\hat{\Delta}$	0.4275 (0.0119) [33.3%]	0.1867 (0.0046) [16.7%]	0.0942 (0.0021) [11.4%]	0.0489 (0.0010) [3.0%]

random matrix F. The entry in parenthesis are the estimated standard deviations of those averaged losses and the PRIAL (or percentage reduction in average loss) relative to the risk of the best multiple estimator are in brackets. Comparisons are made for four different choices of the population matrix Δ between the estimator in (4.1) $\hat{\Delta}_K$ (orthogonally invariant and better than the best multiple) and the orthogonally invariant and minimax estimator of Theorem 4. Our proposed estimator $\hat{\Delta}$ provides substantial risk reduction over the best multiple estimator (as much as 35%) and is markedly better than $\hat{\Delta}_K$ which modifies all eigenvalues in the same direction.

Finally, we remark that the estimators of Konno (1.4) and (4.1) have the same functional form (but are not identical) and are specialized respectively to the loss functions (1.2) and (1.5). Our improved orthogonally invariant minimax estimator $\hat{\Delta}$ in Theorem 4, specialized for the loss (1.5), could also be slightly modified (by another choice of $\mathbf{d}$) to serve under the loss (1.2). The estimator thus obtained can be shown by simulations to yield similar risk reductions over the best multiple estimator (1.3) for the loss (1.2). However, we are unable to establish the minimaxity via unbiased risk estimate methods.

Acknowledgments

The research was supported by the Natural Sciences and Engineering Research Council of Canada. The first author was also supported by FCAR of the province of Québec.

References

[1] Eaton, M. L. (1970). *Some Problems in Covariance Estimation.* Technical Report No. 49. Stanford University.

[2] Fraser, D. A. S. (1976). *Probability and Statistics, Theory and Applications.* DAI Press.

[3] Haff, L. R. (1983). Solutions of the Euler–Lagrange equations for certain multivariate normal estimation problems. Unpublished manuscript.

[4] Haff, L. R. (1988). The variational form of certain Bayes estimators. Unpublished manuscript.

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

[6] Konno, Y. (1989). Improved estimators of eigenvalues of scale matrix of the multivariate F-distribution. University of Tsukuba, Japan. Unpublished manuscript.

[7] Leung, P. L., and Muirhead, R. J. (1987). Estimation of parameter matrices and eigenvalues in Manova and canonical correlation analysis. *Ann. Statist.* **7** 1651–1666.

[8] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

[9] Muirhead, R. J., and Verathaworn, T. (1985). On estimating the latent roots of $\Sigma_1 \Sigma_2^{-1}$. In *Multivariate Analysis VI* (P. R. Krishnaiah, Ed.), pp. 431–447. Academic Press, New York/London.

[10] Perron, F. (1989). *Minimax Estimators of a Covariance Matrix.* Technical Report D.M.S., No. 89–23, Université de Montréal.

[11] Srivastava, M. S., and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics.* North Holland, Amsterdam.

[12] Stein, C. (1975). Estimation of a covariance matrix. Rietz Lecture, 39th Annual Meeting IMS. Atlanta, Georgia.

[13] Takemura, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukuba J. Math.* **8** 367–376.

[14] Verathaworn, T. (1983). *Decision Theoretic Approaches to Estimating Covariance Matrices in the One-Sample and Two-Sample Problems.* Ph.D. thesis. University of Michigan.