The Annals of Statistics
1987, Vol. 15, No. 4, 1651–1666

ESTIMATION OF PARAMETER MATRICES AND EIGENVALUES IN MANOVA AND CANONICAL CORRELATION ANALYSIS

BY PUI LAM LEUNG AND ROBB J. MUIRHEAD[1]

University of California at Los Angeles and University of Michigan

We consider the problem of estimating parameter matrices which occur in the noncentral Wishart, noncentral multivariate F and canonical correlations distributions. A decision-theoretic approach is taken with squared error as the loss function. In these three settings the eigenvalues of the parameter matrices are of primary interest. Sensible estimates of these are obtained by restricting attention to orthogonally invariant estimates of the parameter matrices, whose eigenvalues are functions only of sample eigenvalues.

1. Introduction. Many classical multivariate procedures revolve around random and parameter matrices and their eigenstructures. Invariance and other considerations tend to focus a great deal of attention on the eigenvalues. This paper is concerned with the estimation of parameter matrices and their eigenvalues which occur in the noncentral Wishart, noncentral multivariate F and canonical correlations distributions. The first two of these distributions occur in the standard MANOVA and discriminant analysis settings.

The eigenvalue estimation problem may be stated as follows: Given sample eigenvalues $l_1, \ldots, l_m$ $(l_1 > \cdots > l_m > 0)$, use these to estimate population eigenvalues $\omega_1, \ldots, \omega_m$ $(\omega_1 \geq \cdots \geq \omega_m \geq 0)$. In the situations considered, the usual estimate of ω_i is either l_i or a simple linear function of l_i. Such an estimate, however, ignores information about ω_i in l_j, for $j \neq i$. We are interested here in a decision-theoretic approach to the estimation problems. Ideally, such an approach would specify a loss function in terms of the parameters $\omega_1, \ldots, \omega_m$ and risk calculations would involve expectations of this loss taken with respect to the joint distribution of $l_1, \ldots, l_m$. Unfortunately, this does not seem feasible, due primarily to the complexity of the distribution of the ordered eigenvalues involved [see James (1964) or Muirhead (1982), Sections 10.4 and 11.3.4].

The approach taken in this paper, both in MANOVA and canonical correlations, is to construct a random matrix F whose eigenvalues are $l_1, \ldots, l_m$ and a parameter matrix Δ whose eigenvalues are $\omega_1, \ldots, \omega_m$. In the noncentral Wishart setting the choice of these matrices is obvious: We take F to be the noncentral Wishart matrix and Δ to be the noncentrality matrix. In the two other situations, the choice is not as clear-cut. There are many ways of choosing the matrix

Received May 1986; revised February 1987.

[1]Work supported by the National Science Foundation and the Mathematical Sciences Research Center of the Australian National University.

AMS 1980 *subject classification*. Primary 62H12.

Key words and phrases. Estimation of parameter matrices and eigenvalues, noncentral Wishart distribution, dominance, MANOVA, canonical correlations.

F; it turns out, however, that the distribution theory and resulting risk calculations are greatly simplified by the choice of a nonobservable F (with observable eigenvalues $l_1, \ldots, l_m$). Our approach may then be summarized as follows: Act as if F is observable and use it to construct an "estimate" $\hat{\Delta}(F)$ of Δ, using the squared-error loss function

$$L(\Delta, \hat{\Delta}) = \operatorname{tr}(\hat{\Delta} - \Delta)^2. \tag{1.1}$$

The eigenvalues of $\hat{\Delta}(F)$ may then be regarded as estimates of $\omega_1, \ldots, \omega_m$, the eigenvalues of Δ. We insist that these be *proper* estimates in the sense that they depend on F only through its eigenvalues $l_1, \ldots, l_m$ and, hence, are observable. To this end we consider only orthogonally invariant estimates of Δ which have the same eigenvectors as F and whose eigenvalues are functions only of $l_1, \ldots, l_m$, i.e., estimates of the form

$$\hat{\Delta}(F) = H\phi(L)H', \tag{1.2}$$

where H is an $m \times m$ orthogonal matrix such that $F = HLH'$, with $L = \operatorname{diag}(l_1, \ldots, l_m)$ and $\phi(L) = \operatorname{diag}(\phi_1(L), \ldots, \phi_m(L))$. For $i = 1, \ldots, m$, the observable random variable $\phi_i(L)$ may then be regarded as an estimate of ω_i. In the two situations considered, Δ has an orthogonally invariant unbiased estimate $\hat{\Delta}_U \equiv \alpha F + \beta I$ for certain constants α and β. Orthogonally invariant estimates which dominate $\hat{\Delta}_U$ are found having the form $c_1 F + c_2 I + c_3(\operatorname{tr} F)^{-1} I$ for certain constants c_1, c_2 and c_3; the corresponding estimate of ω_i is then $c_1 l_i + c_2 + c_3(\sum_{j=1}^m l_j)^{-1}$, $i = 1, \ldots, m$.

The preceding discussion has emphasized the nonobservability of the matrix F outside of the noncentral Wishart setting. It should be noted, however, that this problem arises only in the MANOVA and canonical correlations contexts because of certain transformations made using parameter matrices (see Sections 3 and 4). The matrix F is, however, observable in the context of other estimation problems, and the dominance results presented in Sections 3 and 4 pertain to proper estimation problems when viewed in this context.

Eigenvalue estimation problems have been studied in other areas by various authors. Most have considered the problem of estimating a population covariance matrix Σ given a Wishart matrix; restricting attention to orthogonally invariant estimates of the type just described reduces the problem to one of estimating the eigenvalues of Σ. The most relevant papers here are those of Stein (1977), Haff (1980, 1985), Lin and Perlman (1985) and Dey and Srinivasan (1985). Estimation of the eigenvalues of $\Sigma_1\Sigma_2^{-1}$, a two sample problem, has also been studied by Muirhead and Verathaworn (1985) and a start was made in the canonical correlation setting by Muirhead and Leung (1985).

The present paper is organized as follows. The main results associated with the noncentral Wishart distribution are given in Section 2, those in the noncentral multivariate F are given in Section 3 and those in canonical correlation analysis are given in Section 4. The results of Monte Carlo studies also appear in Sections 3 and 4.

2. The noncentral Wishart distribution. Suppose the $n \times m$ matrix Z is a matrix of independent normal variables with unit variance and matrix of means $E(Z) = M$, so that the $m \times m$ matrix $A = Z'Z$ has the noncentral Wishart distribution $A \sim W_m(n, I, \Delta)$, with nonnegative definite noncentrality matrix $\Delta = M'M$ [see, e.g., Muirhead (1982), Section 10.3]. The problem of estimating the mean matrix M in a decision-theoretic context has been considered by various authors, among them Efron and Morris (1972, 1976) and Zidek (1978). Here we are concerned with estimating the noncentrality matrix Δ directly. Given an observation on A, the problem considered is that of estimating Δ by $\hat{\Delta}(A)$ using the loss function (1.1).

Matrices having noncentral Wishart distributions occur in many multivariate testing problems [see Muirhead (1982), Chapter 10]. In particular, the noncentral Wishart distribution is important in a MANOVA situation where the covariance structure of the error matrix is either known or can be estimated accurately.

For example, consider the usual multivariate linear model

$$Y = XB + E,$$

where Y is $N \times m$, X is $N \times p$ of rank p, B is $p \times m$ and the N rows of E have independent identical m-variate normal distributions with mean zero and *known* covariance matrix Σ. Without loss of generality we can take $\Sigma = I$. The likelihood ratio statistic for testing H: $B = 0$ against K: $B \neq 0$ is

$$\Lambda = \exp\left(-\tfrac{1}{2}\operatorname{tr} A\right),$$

where A is the usual regression (or hypothesis) matrix of sums of squares and sums of products given by

$$A = Y'X(X'X)^{-1}X'Y.$$

Here A has the noncentral Wishart distribution $W_m(n, I, \Delta)$, with $n = p$ and $\Delta = B'X'XB$. The power of this test (and others like it) is determined by the noncentrality matrix Δ or, more precisely, by its eigenvalues. In some instances it may be possible to choose among competing tests by estimating Δ (or its eigenvalues) using a preliminary sample. In the univariate setting ($m = 1$) the estimation problem reduces to that of estimating the noncentrality parameter in a noncentral χ^2 distribution, considered earlier by Perlman and Rasmussen (1975), Saxena and Alam (1982), Chow and Hwang (1982) and Chow (1987).

Since $E(A) = nI + \Delta$ [see Muirhead (1982), page 442], it follows that an unbiased estimate of Δ is

$$\hat{\Delta}_U = A - nI. \tag{2.1}$$

Its risk is given in the following lemma.

LEMMA 2.1. *Using the loss function* (1.1), *the risk of* $\hat{\Delta}_U$ *is*

$$R(\hat{\Delta}_U, \Delta) = 2(m+1)\operatorname{tr}\Delta + nm(m+1). \tag{2.2}$$

PROOF. The risk is

$$R(\hat{\Delta}_U, \Delta) = E\left[\operatorname{tr}(\hat{\Delta}_U - \Delta)^2\right] \tag{2.3}$$
$$= E\left[\operatorname{tr}(A^2)\right] - n^2 m - \operatorname{tr}(\Delta^2) - 2n \operatorname{tr} \Delta,$$

where we have used the fact that $E(A) = nI + \Delta$, so that $E(\operatorname{tr} A) = nm + \operatorname{tr} \Delta$. Using the standard vector notation, the first term on the right of (2.3) may be written as

$$E\left[\operatorname{tr}(A^2)\right] = \operatorname{tr} E\left[\operatorname{vec}(A)\operatorname{vec}(A)'\right] \tag{2.4}$$
$$= \operatorname{tr}\{\operatorname{Cov}(\operatorname{vec}(A)) + E\left[\operatorname{vec}(A)\right] E\left[\operatorname{vec}(A)'\right]\}.$$

Now, the covariance matrix of $\operatorname{vec}(A)$ is [Magnus and Neudecker (1979)]

$$\operatorname{Cov}(\operatorname{vec}(A)) = (I_{m^2} + K)(nI_{m^2} + (I_m \otimes \Delta) + (\Delta \otimes I_m)), \tag{2.5}$$

where

$$K = \sum_{i, j=1}^{m} (H_{ij} \otimes H'_{ij}),$$

with H_{ij} being an $m \times m$ matrix with the $i - j$ element equal to 1 and all other elements 0. Using (2.5) and the fact that $E[\operatorname{vec}(A)] = n \operatorname{vec}(I) + \operatorname{vec}(\Delta)$, it follows easily that

$$E\left[\operatorname{tr}(A^2)\right] = \operatorname{tr}(\Delta^2) + 2(n + m + 1)\operatorname{tr} \Delta + nm(n + m + 1) \tag{2.6}$$

and substitution of (2.6) in (2.3) gives the desired result (2.2) and completes the proof. □

It is fairly straightforward to show that no linear estimate of the form $\alpha A + \beta I$ dominates the unbiased estimate $\hat{\Delta}_U$ using the loss function (1.1). Of course, $\hat{\Delta}_U$ need not be nonnegative definite, so is itself inadmissible, being dominated by $\hat{\Delta}_U^+$, a matrix with the same eigenvectors as $\hat{\Delta}_U$ but with any negative eigenvalues of $\hat{\Delta}_U$ being replaced by zero.

We now turn to nonlinear estimates of Δ of the form

$$\hat{\Delta}_\alpha = \hat{\Delta}_U + \frac{\alpha}{\operatorname{tr} A} I. \tag{2.7}$$

In the univariate case ($m = 1$), Perlman and Rasmussen (1975) gave an empirical Bayes justification for such estimates. Such a justification can also be given in the multivariate case, but will not be expanded on here. When $m = 1$ Perlman and Rasmussen (1975) demonstrated that for certain values of α, $\hat{\Delta}_\alpha$ dominates the unbiased estimate $\hat{\Delta}_U$. A similar result holds when $m > 1$ and is given in the following theorem.

THEOREM 2.2. *If* $mn > 4$ *and* $0 < \alpha < [4(mn - 4)]/m$, *then* $\hat{\Delta}_\alpha$ *dominates* $\hat{\Delta}_U$ *for all nonnegative definite* Δ.

PROOF. The difference between the risks of $\hat{\Delta}_U$ and $\hat{\Delta}_\alpha$ is

$$G(\Delta) \equiv R(\hat{\Delta}_U, \Delta) - R(\hat{\Delta}_\alpha, \Delta)$$

$$= 2\alpha(mn + \operatorname{tr}\Delta)E\left[\frac{1}{\operatorname{tr}A}\right] - m\alpha^2 E\left[\frac{1}{(\operatorname{tr}A)^2}\right] - 2\alpha. \tag{2.8}$$

It is easily seen that $\operatorname{tr}A$ has the noncentral χ^2 distribution $\chi^2_{mn}(\delta)$, where the noncentrality parameter is $\delta = \operatorname{tr}\Delta$. Using the representation for this distribution as a Poisson mixture of central χ^2 distributions, it follows that

$$E\left[\frac{1}{\operatorname{tr}A}\right] = E\left[\frac{1}{mn + 2K - 2}\right] \tag{2.9}$$

and

$$E\left[\frac{1}{(\operatorname{tr}A)^2}\right] = E\left[\frac{1}{(mn + 2K - 2)(mn + 2K - 4)}\right], \tag{2.10}$$

where K has a Poisson distribution with mean $\frac{1}{2}\operatorname{tr}\Delta$. With the help of (2.9) and (2.10), $G(\Delta)$ may then be written as

$$G(\Delta) = 2\alpha E\left[\frac{\operatorname{tr}\Delta - 2K + 2}{mn + 2K - 2} - \frac{m\alpha}{2(mn + 2K - 2)(mn + 2K - 4)}\right]. \tag{2.11}$$

Since the covariance between $\operatorname{tr}\Delta - 2K$ and $(mn + 2K - 2)^{-1}$ is nonnegative it follows that

$$E\left[\frac{\operatorname{tr}\Delta - 2K}{mn + 2K - 2}\right] \geq$$

and, hence,

$$\begin{aligned} G(\Delta) &\geq \alpha E\left[\frac{4}{mn + 2K - 2} - \frac{m\alpha}{(mn + 2K - 2)(mn + 2K - 4)}\right] \\ &= \alpha E\left[\frac{4mn + 8K - 16 - m\alpha}{(mn + 2K - 2)(mn + 2K - 4)}\right] \\ &\geq \alpha E\left[\frac{4mn - 16 - m\alpha}{(mn + 2K - 2)(mn + 2K - 4)}\right] \end{aligned} \tag{2.12}$$

and this is positive when $mn > 4$ and $0 < \alpha < [4(mn - 4)]/m$. □

A reasonable way of choosing α is by maximizing the lower bound for $G(\Delta)$ in (2.12). The maximizing value is $\alpha = [2(mn - 4)]/m$ and this value satisfies the inequality in Theorem 2.2. The corresponding nonlinear estimate is

$$\hat{\Delta}_{NL} = \hat{\Delta}_U + \frac{2(mn - 4)}{m \operatorname{tr}A} I. \tag{2.13}$$

This dominates $\hat{\Delta}_U$ and, of course, is dominated in turn by $\hat{\Delta}^+_{NL}$, its truncated version where any negative eigenvalues of $\hat{\Delta}_{NL}$ are replaced by zero.

For fixed m and n, the difference between the risks of $\hat{\Delta}_U$ and $\hat{\Delta}_{NL}$ is small if $\operatorname{tr}\Delta$ is large. For fixed m and Δ the difference is about $4/m$ for large n. In fact, it is easily seen from (2.11) that, when $\Delta = 0$, the difference is exactly $[4(mn - 4)]/[m(mn - 2)]$, with a larger value being obtained by truncation.

A Monte Carlo study was carried out to compare these estimates. The results indicated that $\hat{\Delta}_{NL}^{+}$ represents a substantial improvement over $\hat{\Delta}_U$ when $\operatorname{tr}\Delta$ is small or moderate. The differences between $\hat{\Delta}_{NL}^{+}$ and $\hat{\Delta}_U^{+}$ are much less significant.

3. The noncentral multivariate *F*.

3.1. *Discussion.* The preceding section considered an estimation problem applicable in MANOVA when the covariance structure of the error matrix is known. It is more common that this is not known, however, and we now consider this situation. In the typical MANOVA setting, independent $m \times m$ matrices S_1 and S_2 are observed, where S_1 has a noncentral Wishart distribution with n_1 degrees of freedom, covariance matrix Σ and nonnegative definite noncentrality matrix Ω, $S_1 \sim W_m(n_1, \Sigma, \Omega)$, and S_2 has a central Wishart distribution, $S_2 \sim W_m(n, \Sigma)$. We will assume here that $n_1 \geq m$ and $n_2 \geq m$, so that both distributions are nonsingular. A problem of great interest is that of estimating the eigenvalues $\omega_1, \omega_2, \dots, \omega_m$ ($\omega_1 \geq \omega_2 \geq \cdots \geq \omega_m \geq 0$) of the noncentrality matrix Ω. These eigenvalues are important in the problem of testing H: $\Omega = 0$ against K: $\Omega \neq 0$ because they form maximal invariants under a natural group of transformations leaving the testing problem invariant. Any invariant test depends only on $l_1, l_2, \dots, l_m$ ($l_1 > l_2 > \cdots > l_m > 0$), 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 population eigenvalues also play a major role in discriminant analysis. Discussions of MANOVA and discriminant analysis may be found in, e.g., Anderson [(1985), Chapter 8] and Muirhead [(1982), Chapter 10]. In the univariate setting ($m = 1$), the estimation problem reduces to that of estimating the noncentrality parameter in a noncentral F distribution, a problem considered by Perlman and Rasmussen (1975).

It is convenient to transform S_1 and S_2 in a way that greatly simplifies the relevant distribution theory and the resulting risk calculations. 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 $B \sim W_m(n_2, I)$. Note that $\omega_1, \dots, \omega_m$, the eigenvalues of Ω, are also the eigenvalues of the new noncentrality matrix Δ and that $l_1, \dots, l_m$, the eigenvalues of $S_1 S_2^{-1}$, are also the eigenvalues of the positive definite random matrix F defined as

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

We remark that, although F is not observable unless Σ is known, its eigenvalues are observable. We overcome this difficulty using the approach outlined in Section 1, i.e., we estimate Δ by $\hat{\Delta}(F)$ using the squared-error loss function (1.1) and an orthogonally invariant estimate of the form (1.2), so that $\phi_1(L), \dots, \phi_m(L)$, the eigenvalues of $\hat{\Delta}(F)$, are observable and may be regarded as estimates of

$\omega_1, \ldots, \omega_m$. Note that the observability problem arises because of transformations made within the MANOVA context. An equivalent, well-posed problem where it has no bearing is the following: Given an observation F from the distribution of $A^{1/2}B^{-1}A^{1/2}$, where $A \sim W_m(n_1, I, \Delta)$, $B \sim W_m(n_2, I)$ and A and B are independent, estimate the noncentrality matrix Δ. The dominance results in the next section pertain to a proper estimation problem when viewed in this context.

The expectation of F is

$$E(F) = E(A^{1/2}B^{-1}A^{1/2}) = E\left[E(A^{1/2}B^{-1}A^{1/2}|A)\right]$$
$$= \frac{1}{n_2 - m - 1}E(A) = \frac{1}{n_2 - m - 1}(n_1 I + \Delta)$$

(assuming now that $n_2 > m + 1$), so that an unbiased estimate of Δ is the orthogonally invariant estimate $\hat{\Delta}_U$ given by

$$\hat{\Delta}_U = (n_2 - m - 1)F - n_1 I. \tag{3.1}$$

The corresponding estimate of ω_i derived from $\hat{\Delta}_U$ is thus $(n_2 - m - 1)l_i - n_1$. The only orthogonally invariant estimates considered in this paper are ones which dominate $\hat{\Delta}_U$ and have the form $c_1 F + c_2 I + c_3(\operatorname{tr} F)^{-1}$ for certain constants c_1, c_2 and c_3.

3.2. *Main results*. When $m = 1$, Perlman and Rasmussen (1975) showed that, for certain values of α, the linear estimate $\alpha\hat{\Delta}_U$ dominates the unbiased estimate $\hat{\Delta}_U$ with respect to squared-error loss. A similar result holds when $m > 1$ and the loss function (1.1) is used, and is given in the following theorem. We assume, throughout this section and the next, that $n_2 > m + 3$.

THEOREM 3.1. *The estimate $\alpha\hat{\Delta}_U$ dominates $\hat{\Delta}_U$ provided that*

$$\max\left(0, \frac{n_2 - m - 5}{n_2 - m - 1}\right) \leq \alpha < 1.$$

It is seen from the proof in Section 3.3 that an optimal value of α is α^* given by

$$\alpha^* = \frac{n_2 - m - 3}{n_2 - m - 1}, \tag{3.2}$$

assuming $n_2 > m + 3$. The corresponding linear estimate $\hat{\Delta}_L$ given by

$$\begin{aligned} \hat{\Delta}_L &= \frac{n_2 - m - 3}{n_2 - m - 1}\hat{\Delta}_U \\ &= (n_2 - m - 3)F - \frac{n_1(n_2 - m - 3)}{n_2 - m - 1}I \end{aligned} \tag{3.3}$$

thus dominates $\hat{\Delta}_U$.

Two remarks are worth making. First, a consequence of Theorem 3.1 is that, if $n_2 = m + 4$ or $n_2 = m + 5$, $\hat{\Delta}_U$ is dominated by the trivial estimate $\hat{\Delta} \equiv 0$. Second, $\hat{\Delta}_L$ need not be nonnegative definite, so is itself inadmissible, being dominated by $\hat{\Delta}_L^+$, a truncated version of $\hat{\Delta}_L$ which has the same eigenvectors but with any negative eigenvalues being replaced by zero. It also turns out that $\hat{\Delta}_L$ is dominated by nonlinear estimates of the form

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

This result is a consequence of the following theorem, which is a multivariate generalization of a result of Perlman and Rasmussen (1975).

THEOREM 3.2. *The estimate* $\hat{\Delta}_{\alpha,\beta}$ *given by* (3.4) *dominates* $\alpha\hat{\Delta}_U$ *provided* $mn_1 > 4$,

$$0 < \alpha \le 1 + \frac{2}{m(n_2 - m - 1)},$$

and

$$0 < \beta < \frac{4\alpha(n_1 + n_2 - m - 1)(mn_1 - 4)}{m(n_2 - m + 3)(n_2 - m + 1)}.$$

If we take the value α^* given by (3.2) for α (which corresponds to $\hat{\Delta}_L$), then the proof of Theorem 3.2 in Section 3.3 shows that an optimal value for β is β^* given by

$$\beta^* = \frac{2(n_2 - m - 3)(n_1 + n_2 - m - 1)(mn_1 - 4)}{m(n_2 - m - 1)(n_2 - m + 3)(n_2 - m + 1)}. \tag{3.5}$$

The corresponding nonlinear estimate $\hat{\Delta}_{NL} \equiv \hat{\Delta}_{\alpha^*,\beta^*}$ thus dominates $\hat{\Delta}_L$. It is also the case that $\hat{\Delta}_{NL}$ is dominated in turn by its truncated version $\hat{\Delta}_{NL}^+$.

A Monte Carlo study was carried out to compare $\hat{\Delta}_L^+$ and $\hat{\Delta}_{NL}^+$ with $\hat{\Delta}_U^+$, the truncated version of the unbiased estimate $\hat{\Delta}_U$. The results indicate that both $\hat{\Delta}_L^+$ and $\hat{\Delta}_{NL}^+$ represent substantial improvements over $\hat{\Delta}_U^+$, particularly when n_1 and n_2 are small. The study also revealed that, although $\hat{\Delta}_{NL}$ dominates $\hat{\Delta}_L$, $\hat{\Delta}_L^+$ has a tendency to improve on $\hat{\Delta}_{NL}^+$. In view of this, there seems little to be gained by using $\hat{\Delta}_{NL}^+$ in preference to $\hat{\Delta}_L^+$.

3.3. *Proofs.* To prove Theorem 3.1, we compute the risks of $\hat{\Delta}_U$ and $\alpha\hat{\Delta}_U$. For this we need the expectation of $\operatorname{tr}(F^2)$. This is given in the following lemma, whose proof is reasonably straightforward and is omitted.

LEMMA 3.3.

$$E\left[\operatorname{tr}(F^2)\right] = \beta_0\left[(\operatorname{tr}\Delta)^2 + \beta_1\operatorname{tr}(\Delta^2) + \beta_2\operatorname{tr}\Delta + \beta_3\right], \tag{3.6}$$

where

$$\begin{aligned}\beta_0 &= \frac{1}{(n_2 - m)(n_2 - m - 1)(n_2 - m - 3)},\\ \beta_1 &= n_2 - m - 1,\\ \beta_2 &= 2[(n_2 - m)(n_1 + m + 1) + (m - 1)(n_1 - 1)],\\ \beta_3 &= \frac{mn_1\beta_2}{2}.\end{aligned} \tag{3.7}$$

PROOF OF THEOREM 3.1. The difference between the risks of $\hat{\Delta}_U$ and $\alpha\hat{\Delta}_U$ is

$$G(\Delta) \equiv E\left[\mathrm{tr}(\hat{\Delta}_U - \Delta)^2\right] - E\left[\mathrm{tr}(\alpha\hat{\Delta}_U - \Delta)^2\right],$$

which, on using the result of Lemma 3.3, may be expressed as

$$\begin{aligned}G(\Delta) &= a(1 - \alpha^2)(\mathrm{tr}\,\Delta)^2 + (1 - \alpha)[b(1 + \alpha) - 1 + \alpha]\mathrm{tr}(\Delta^2)\\ &\quad + c(1 - \alpha^2)\mathrm{tr}\,\Delta + d(1 - \alpha^2)\\ &= (1 - \alpha)[(a + b)(1 + \alpha) - 1 + \alpha]\mathrm{tr}(\Delta^2)\\ &\quad + 2a(1 - \alpha^2)\sum_{i<j}^{m} \omega_i\omega_j + c(1 - \alpha^2)\mathrm{tr}\,\Delta + d(1 - \alpha^2),\end{aligned} \tag{3.8}$$

where

$$\begin{aligned}a &= \frac{n_2 - m - 1}{(n_2 - m)(n_2 - m - 3)},\\ b &= \frac{n_2 - m + 1}{(n_2 - m)(n_2 - m - 3)},\\ c &= \frac{2(n_1 + n_2 - m - 1)[(m + 1)(n_2 - m) - (m - 1)]}{(n_2 - m)(n_2 - m - 3)},\end{aligned}$$

and

$$d = \frac{mn_1c}{2}.$$

Under the conditions stated in Theorem 3.1, the constant term and the coefficients of $\mathrm{tr}\,\Delta$ and $\Sigma_{i<j}^{m}\omega_i\omega_j$ are positive. The proof is completed by noting that the coefficient of $\mathrm{tr}(\Delta^2)$ is nonnegative provided

$$\alpha \geq \frac{1 - a - b}{1 + a + b} = \frac{n_2 - m - 5}{n_2 - m - 1}. \qquad \square$$

It is seen from (3.8) that

$$G(\Delta) > (1 - \alpha)[(a + b)(1 + \alpha) - 1 + \alpha]\mathrm{tr}(\Delta^2);$$

the optimal value α^* given by (3.2) is the value of α that maximizes this lower bound for $G(\Delta)$.

We turn now to the proof of Theorem 3.2. In this, bounds are used for the first two moments of $(\operatorname{tr} F)^{-1}$ which are expressed in terms of expectations involving a Poisson random variable K. They are consequences of the following lemma.

LEMMA 3.4. *Let K be a Poisson random variable with mean $\frac{1}{2}\operatorname{tr}\Delta$. Then, if $mn_1 > 2$,*

$$\frac{1}{m}E\left[\frac{2+m(n_2-m-1)}{mn_1+2K-2}\right] \le E\left[\frac{1}{\operatorname{tr} F}\right] \le E\left[\frac{n_2-m+1}{mn_1+2K-2}\right] \tag{3.9}$$

and, if $mn_1 > 4$,

$$\begin{aligned}&\frac{1}{m^2}E\left[\frac{(2+m(n_2-m-1))(4+m(n_2-m-1))}{(mn_1+2K-2)(mn_1+2K-4)}\right]\\ &\quad\le E\left[\frac{1}{(\operatorname{tr} F)^2}\right]\\ &\quad\le E\left[\frac{(n_2-m+3)(n_2-m+1)}{(mn_1+2K-2)(mn_1+2K-4)}\right].\end{aligned} \tag{3.10}$$

PROOF. To prove (3.9) we begin by conditioning on A. Given A, the conditional distribution of F^{-1} is $W_m(n_2, A^{-1})$ and the Wishart identity [see Haff (1980), Equation (2.4)] may be used to show that

$$E\left[\frac{1}{\operatorname{tr} F}\middle|A\right] = \frac{1}{\operatorname{tr} A}\left\{2E\left[\frac{\operatorname{tr} F^2}{(\operatorname{tr} F)^2}\middle|A\right] + (n_2-m-1)\right\}. \tag{3.11}$$

Using the fact that [see, e.g., Haff (1980), Lemma 5.2]

$$\frac{1}{m} \le \frac{\operatorname{tr} F^2}{(\operatorname{tr} F)^2} \le 1, \tag{3.12}$$

it follows from (3.11) that

$$\frac{2+m(n_2-m-1)}{m\operatorname{tr} A} \le E\left[\frac{1}{\operatorname{tr} F}\middle|A\right] \le \frac{n_2-m-1}{\operatorname{tr} A} \tag{3.13}$$

and the desired result now follows on taking expectations with respect to the $W_m(n_1, I, \Delta)$ distribution of A and using (2.9). To prove (3.10), the Wishart identity may be used to show that

$$E\left[\frac{1}{(\operatorname{tr} F)^2}\middle|A\right] = \frac{1}{\operatorname{tr} A}\left\{4E\left[\frac{\operatorname{tr} F^2}{(\operatorname{tr} F)^3}\middle|A\right] + (n_2-m-1)E\left[\frac{1}{\operatorname{tr} F}\middle|A\right]\right\}. \tag{3.14}$$

The inequality (3.12) guarantees that

$$\frac{1}{m}E\left[\frac{1}{\operatorname{tr} F}\middle|A\right] \le E\left[\frac{\operatorname{tr} F^2}{(\operatorname{tr} F)^3}\middle|A\right] \le E\left[\frac{1}{\operatorname{tr} F}\middle|A\right],$$

and using these bounds for $E[\operatorname{tr} F^2/(\operatorname{tr} F)^3|A]$ in (3.14) shows that

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

and hence, using (3.13), that

$$\begin{aligned}\frac{[4+m(n_2-m-1)][2+m(n_2-m-1)]}{m^2(\operatorname{tr} A)^2} &\le E\left[\frac{1}{(\operatorname{tr} F)^2}\middle|A\right]\\ &\le \frac{(n_2-m+3)(n_2-m+1)}{(\operatorname{tr} A)^2}.\end{aligned}$$

The desired result (3.10) is now immediate on taking expectations with respect to A and using (2.10). □

It is worth pointing out that the upper and lower bounds in Lemma 3.4 are identical when $m = 1$. The resulting identities were utilized in the univariate setting by Perlman and Rasmussen (1975).

PROOF OF THEOREM 3.2. The difference between the risks of $\alpha\hat{\Delta}_U$ and $\hat{\Delta}_{\alpha,\beta}$ is easily seen to be

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

Using the upper bound for $E[1/\operatorname{tr} F]$ and the lower bound for $E[1/(\operatorname{tr} F)^2]$ given in Lemma 3.4, it follows that

$$\begin{aligned}H(\Delta) \ge{}& 2\beta(\alpha m n_1+\operatorname{tr}\Delta)\left(\frac{2}{m}+n_2-m-1\right)E\left[\frac{1}{mn_1+2K-2}\right]\\ &-m\beta^2(n_2-m+3)(n_2-m+1)\\ &\quad\times E\left[\frac{1}{(mn_1+2K-2)(mn_1+2K-4)}\right]\\ &-2\alpha\beta(n_2-m-1)\\ ={}& 2\beta E\left[\{(2/m+n_2-m-1)\operatorname{tr}\Delta - 2\alpha K(n_2-m-1)\right.\\ &\qquad\left.+2\alpha(n_1+n_2-m-1)\}/\{mn_1+2K-2\}\right]\\ &-m\beta^2 E\left[\frac{(n_2-m+3)(n_2-m+1)}{(mn_1+2K-2)(mn_1+2K-4)}\right],\end{aligned}$$

where K has a Poisson distribution with mean $\frac{1}{2}\operatorname{tr}\Delta$.

For $\alpha > 0$, the covariance between $(mn_1 + 2K - 2)^{-1}$ and $(2/m + n_2 + m - 1)\mathrm{tr}\,\Delta - 2\alpha K(n_2 - m - 1)$ is nonnegative and, hence,

$$E\left[\frac{(2/m + n_2 - m - 1)\mathrm{tr}\,\Delta - 2\alpha K(n_2 - m - 1)}{mn_1 + 2K - 2}\right]$$
$$\geq \left[\frac{2}{m} + (1 - \alpha)(n_2 - m - 1)\right]\mathrm{tr}\,\Delta E\left[\frac{1}{mn_1 + 2K - 2}\right] \geq 0,$$

provided $0 < \alpha \leq 1 + 2/[m(n_2 - m - 1)]$. It then follows that

$$(3.15)\qquad \begin{aligned} H(\Delta) &\geq \beta E\left[\frac{4\alpha(n_1 + n_2 - m - 1)}{mn_1 + 2K - 2} - \frac{m\beta(n_2 - m + 3)(n_2 - m + 1)}{(mn_1 + 2K - 2)(mn_1 + 2K - 4)}\right] \\ &\geq \beta E\left[\frac{4\alpha(n_1 + n_2 - m - 1)(mn_1 - 4) - m\beta(n_2 - m + 3)(n_2 - m + 1)}{(mn_1 + 2K - 2)(mn_1 + 2K - 4)}\right], \end{aligned}$$

and the proof is completed by noting that the right side is positive when $mn_1 > 4$ and

$$0 < \beta < \frac{4\alpha(n_1 + n_2 - m - 1)(mn_1 - 4)}{m(n_2 - m + 3)(n_2 - m + 1)}. \qquad \square$$

When α takes the value α^* given by (3.2), the optimal value β^* given by (3.5) is the value of β maximizing the lower bound for $H(\Delta)$ given in (3.15).

4. Canonical correlation coefficients.

4.1. *Discussion and main results.* Suppose the $(p + q) \times (p + q)$ positive definite matrix S has the Wishart $W_{p+q}(n, \Sigma)$ distribution, $n \geq p + q$, and partition S and Σ as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where S_{11} and Σ_{11} are $p \times p$ and S_{22} and Σ_{22} are $q \times q$, with $p \leq q$. The population canonical correlation coefficients are $\rho_1, \ldots, \rho_p$ $(1 \geq \rho_1 \geq \cdots \geq \rho_p \geq 0)$, where $\rho_1^2, \ldots, \rho_p^2$ are the eigenvalues of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. The positive square roots $r_1, \ldots, r_p$ $(1 > r_1 > \cdots > r_p > 0)$ of $r_1^2, \ldots, r_p^2$, the eigenvalues of $S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}$, are the sample canonical correlation coefficients. Discussions of canonical correlation analysis may be found in, e.g., Anderson [(1985), Chapter 12] and Muirhead [(1982), Chapter 11]. These eigenvalues are also important in the problem of testing independence of two sets of jointly normally distributed variables, i.e., in testing H: $\Sigma_{12} = 0$ against K: $\Sigma_{12} \neq 0$, as they form maximal invariants under a natural group of transformations leaving the testing problem invariant. Any invariant test statistic is a function of $r_1^2, \dots, r_p^2$ and has a power function which depends on Σ only through $\rho_1^2, \dots, \rho_p^2$.

In an earlier paper, Muirhead and Leung (1985) considered the problem of estimating the parameter $\omega_i = \rho_i^2/(1-\rho_i^2)$ using linear functions of $l_i = r_i^2/(1-r_i^2)$ $(i = 1, \dots, p)$. Here we extend this work by considering certain nonlinear estimates. It is worth noting that, when estimating correlation coefficients, the choice of scale is an important consideration. Transforming from $r_1, \dots, r_p$ to $l_1, \dots, l_p$ changes the support from $1 > r_1 > r_2 > \cdots > r_p > 0$, where the distributions can be highly skewed, to $\infty > l_1 > l_2 > \cdots > l_p > 0$, with more symmetric distributions and should allow more stable estimation in view of the fact that the transformed parameters ω_i act much more like location parameters. It is interesting to note that Fisher (1915) first suggested making such a transformation in connection with an ordinary correlation coefficient.

The random variables $l_1, \dots, l_p$ are the eigenvalues of the random matrix

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

where $A = \Sigma_{11\cdot 2}^{-1/2}S_{12}S_{22}^{-1}S_{21}\Sigma_{11\cdot 2}^{-1/2}$ and $B = \Sigma_{11\cdot 2}^{-1/2}S_{11\cdot 2}\Sigma_{11\cdot 2}^{-1/2}$, with $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and $S_{11\cdot 2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$. Put $X = \Sigma_{22}^{-1/2}S_{22}\Sigma_{22}^{-1/2}$; from standard distribution theory [see, e.g., Muirhead (1985), Theorem 3.2.10 and Section 10.3], it follows that $B \sim W_p(n-q, I_p)$ and is independent of X and A, that $X \sim W_q(n, I_q)$ and that, conditional on X, the distribution of A is noncentral Wishart $W_p(q, I_p, \Omega)$, where the noncentrality matrix Ω is

$$\Omega = \Sigma_{11\cdot 2}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}X\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11\cdot 2}^{-1/2}.$$

It is not difficult to show that the distribution of F depends on Σ only through the parameter matrix Δ given by

$$\Delta = \Sigma_{11\cdot 2}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11\cdot 2}^{-1/2}, \tag{4.1}$$

which has as its eigenvalues the parameters we wish to estimate, namely $\omega_i = \rho_i^2/(1-\rho_i^2)$, $i = 1, \dots, p$. Note that, as in the MANOVA situation considered in Section 3, we have cast the estimation problem in terms of a random matrix F which is not observable unless $\Sigma_{11\cdot 2}$ is known. However, the eigenvalues $l_1, \dots, l_p$ of this matrix are observable. Exactly as in Section 3, we overcome the nonobservability difficulty by estimating Δ by $\hat{\Delta}(F)$ using the loss function (1.1) and an orthogonally invariant estimate of the form (1.2), so that the eigenvalues of $\hat{\Delta}(F)$, which are observable, may be regarded as estimates of $\omega_1, \dots, \omega_p$.

It is shown in Muirhead and Leung (1985) (and easily verified directly) that

$$E(F) = \frac{1}{n-p-q-1}(n\Delta + qI_p),$$

(assuming now that $n > p + q + 1$) so that an unbiased estimate of Δ is

$$\hat{\Delta}_U = \frac{n-p-q-1}{n}F - \frac{q}{n}I_p. \tag{4.2}$$

This is dominated by estimates of the form $\alpha\hat{\Delta}_U$. The following theorem is a slight modification of Theorem 3 in Muirhead and Leung (1985).

THEOREM 4.1. *The estimate* $\alpha\hat{\Delta}_U$ *dominates* $\hat{\Delta}_U$ *provided* $n - p - q - 1 > 0$ *and*

$$\max\left(0, \frac{n^2 - n(p+q+7) + 2(p+q+1)}{(n+2)(n-p-q-1)}\right) \le \alpha < 1.$$

An optimal value of α is

$$\alpha^* = \frac{n(n-p-q-3)}{(n+2)(n-p-q-1)}, \tag{4.3}$$

assuming $n > p + q + 3$. (This value of α maximizes a lower bound for the difference between the risks of $\hat{\Delta}_U$ and $\alpha\hat{\Delta}_U$.) The corresponding linear estimate

$$\begin{aligned}\hat{\Delta}_L &= \frac{n(n-p-q-3)}{(n+2)(n-p-q-1)}\hat{\Delta}_U \\ &= \frac{n-p-q-1}{n+2}F - \frac{q(n-p-q-3)}{(n+2)(n-p-q-1)}I_p\end{aligned} \tag{4.4}$$

thus dominates $\hat{\Delta}_U$. Comments similar to those made after Theorem 3.1 are applicable here as well, and will not be repeated.

We turn now to nonlinear estimates of the form

$$\hat{\Delta}_{\alpha,\beta} = \alpha\hat{\Delta}_U + \frac{\beta}{\operatorname{tr} F}I. \tag{4.5}$$

When $p = 1$ (the multiple correlation setting), Muirhead (1985) showed that, for certain values of α and β, $\hat{\Delta}_{\alpha,\beta}$ dominates $\alpha\hat{\Delta}_U$. A similar result holds when $p > 1$ as the following theorem, whose proof follows, shows.

THEOREM 4.2. *The estimate* $\hat{\Delta}_{\alpha,\beta}$ *given by* (4.5) *dominates* $\alpha\hat{\Delta}_U$ *provided* $pq > 4$,

$$0 < \alpha \le 1 + \frac{2}{p(n-p-q-1)},$$

and

$$0 < \beta < \frac{4\alpha(n-p-1)(pq-4)}{np(n-p-q+3)(n-p-q+1)}.$$

Taking the value α^* given by (4.3) for α, the proof of Theorem 4.2 shows that an optimal value for β is β^* given by

$$\beta^* = \frac{2(n-p-q-3)(n-p-1)(pq-4)}{p(n+2)(n-p-q-1)(n-p-q+3)(n-p-q+1)}. \tag{4.6}$$

The corresponding nonlinear estimate $\hat{\Delta}_{NL} \equiv \hat{\Delta}_{\alpha^*,\beta^*}$ thus dominates $\hat{\Delta}_L$ and, of course, is dominated in turn by its truncated version $\hat{\Delta}_{NL}^+$.

A Monte Carlo study was carried out to compare $\hat{\Delta}_L^+$ and $\hat{\Delta}_{NL}^+$ with $\hat{\Delta}_U^+$. The results indicated that both give substantial gains over $\hat{\Delta}_U^+$, especially when n is small or moderate. The differences between $\hat{\Delta}_L^+$ and $\hat{\Delta}_{NL}^+$ are slight, however.

Proof of Theorem 4.2. The difference between the risks of $\alpha\hat{\Delta}_U$ and $\hat{\Delta}_{\alpha,\beta}$ is

$$J(\Delta) \equiv E\left[\operatorname{tr}\left(\alpha\hat{\Delta}_U - \Delta\right)^2\right] - E\left[\operatorname{tr}\left(\hat{\Delta}_{\alpha,\beta} - \Delta\right)^2\right]$$
$$= \frac{2\beta}{n}(\alpha pq + n\operatorname{tr}\Delta)E\left[\frac{1}{\operatorname{tr}F}\right] - p\beta^2 E\left[\frac{1}{(\operatorname{tr}F)^2}\right] - \frac{2\alpha\beta}{n}(n-p-q-1).$$

Using an almost identical argument to that used in the proof of Theorem 3.2 it may be shown that

$$J(\Delta) \geq \frac{2\beta}{n}E\left[\frac{(2/p+n-p-q-1)n\operatorname{tr}\Delta - 2\alpha K(n-p-q-1) + 2\alpha(n-p-1)}{pq+2K-2}\right]$$
$$-p\beta^2 E\left[\frac{(n-p-q+3)(n-p-q+1)}{(pq+2K-2)(pq+2K-4)}\right]$$

with the expectations on the right being taken with respect to the joint distribution of K and X, where X has the $W_q(n, I_q)$ distribution and where, conditional on X, K has a Poisson distribution with mean $\frac{1}{2}\operatorname{tr}\Omega = \frac{1}{2}\operatorname{tr}\Sigma_{11\cdot 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}X\Sigma_{22}^{-1/2}\Sigma_{21}$. Note that $E(K) = \frac{1}{2}n\operatorname{tr}\Delta$, where Δ is given by (4.1). Since the covariance between $(pq+2K-2)^{-1}$ and $(2/q+n-p-q-1)n\operatorname{tr}\Delta - 2\alpha K(n-p-q-1)$ is nonnegative when $\alpha > 0$, it follows that

$$E\left[\frac{(2/p+n-p-q-1)n\operatorname{tr}\Delta - 2\alpha K(n-p-q-1)}{pq+2K-2}\right]$$
$$\geq [2/p + (1-\alpha)(n-p-q-1)]n\operatorname{tr}\Delta E\left[\frac{1}{pq+2K-2}\right] \geq 0,$$

provided $0 < \alpha \leq 1 + 2/[p(n-p-q-1)]$. It then follows that

$$J(\Delta) \geq \frac{\beta}{n}E\left[\frac{4\alpha(n-p-1)}{pq+2K-2} - \frac{np\beta(n-p-q+3)(n-p-q+1)}{(pq+2K-2)(pq+2K-4)}\right] \tag{4.7}$$
$$\geq \frac{\beta}{n}E\left[\frac{4\alpha(n-p-1)(pq-4) - np\beta(n-p-q+3)(n-p-q+1)}{(pq+2K-2)(pq+2K-4)}\right]$$

and the proof is completed by noting that the right side is positive when $pq > 4$ and

$$0 < \beta < \frac{4\alpha(n-p-1)(pq-4)}{np(n-p-q+3)(n-p-q+1)}. \qquad \square$$

With α taking the value α^* given by (4.3), the optimal value β^* given by (4.6) is the value of β maximizing the lower bound for $J(\Delta)$ given in (4.7).

Acknowledgments. The authors would like to thank L. R. Haff for suggesting the proof of inequality (3.13) based on the Wishart identity. They are also indebted to the referees for some constructive suggestions.

REFERENCES

ANDERSON, T. W (1985). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

CHOW, M. S. (1987). A complete class theorem for estimating a noncentrality parameter. *Ann. Statist.* **15** 800–804.

CHOW, M. S. and HWANG, J. T. (1982). The comparison of estimators for the noncentrality of a chi-square distribution. Technical Report, Dept. Mathematics, Cornell Univ.

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

EFRON, B. and MORRIS, C. (1972). Empirical Bayes on vector observations—an extension of Stein's rule. *Biometrika* **59** 335–347.

EFRON, B. and MORRIS, C. (1976). Multivariate empirical Bayes and estimation of covariance matrices. *Ann. Statist.* 4 22–32.

FISHER, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10** 507–521.

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

HAFF, L. R. (1985). Solutions of the Euler–Lagrange equations for certain multivariate normal estimation problems. Unpublished.

JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475–501.

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

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*. Wiley, New York.

MUIRHEAD, R. J. (1985). Estimating a particular function of the multiple correlation coefficient. *J. Amer. Statist. Assoc.* **80** 923–925.

MUIRHEAD, R. J. and LEUNG, P. L. (1985). Estimating functions of canonical correlation coefficients. *Linear Algebra Appl.* **70** 173–183.

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.) 431–447. North-Holland, Amsterdam.

PERLMAN, M. D. and RASMUSSEN, U. A. (1975). Some remarks on estimating a noncentrality parameter. *Comm. Statist.* 4 455–468.

SAXENA, K. M. L. and ALAM, K. (1982). Estimation of the noncentrality parameter of a chi-squared distribution. *Ann. Statist.* **10** 1012–1016

STEIN, C. (1977). Lectures on multivariate estimation theory. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **74** 4–65, 146, 148 (in Russian).

ZIDEK, J. (1978). Deriving unbiased risk estimators of multinomial mean and regression coefficient estimators using zonal polynomials. *Ann. Statist.* **6** 769–782.

DEPARTMENT OF BIOMATHEMATICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90024

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109