A CLASS OF MINIMAX ESTIMATORS OF A NORMAL MEAN VECTOR FOR ARBITRARY QUADRATIC LOSS AND UNKNOWN COVARIANCE MATRIX

by

James Berger[1] and L. R. Haff[2]
Purdue University   University of California, San Diego

Mimeograph Series #81-7

April 1981

[1]Research supported by the Alfred P. Sloan Foundation and by NSF Grant #MCS 78-02300.

[2]Research supported by NSF Grant #MCS 78-01942 and NIH Grant #CA26666-01.

Short title: Minimax Estimation

AMS 1970 subject classifications: Primary 62C99; Secondary 62F10, 62H99, 62H05.

Key words and phrases. Minimax, normal mean, quadratic loss, unknown covariance matrix, Wishart identities, stochastic ordering of Wishart characteristic roots.

Abstract

Let X be an observation from a p-variate normal distribution ($p \geq 3$) with unknown mean vector θ and unknown positive definite covariance matrix Σ. It is desired to estimate θ under the quadratic loss $L(\delta,\theta,\Sigma) = (\delta-\theta)^t Q(\delta-\theta)/tr(Q\Sigma)$, where Q is a known positive definite matrix. A random matrix W with a Wishart (Σ,n) distribution is assumed to be available, from which Σ can be estimated. A broad class of minimax estimators of θ is developed, using a novel combination of techniques developed separately for estimating means and covariance matrices. These minimax estimators can offer significant improvement upon the usual estimator, $\delta^0(X) = X$, of θ.

1. Introduction

Assume $X = (X_1,\ldots,X_p)^t$ is a p-dimensional random vector ($p \geq 3$) which is normally distributed with mean vector $\theta = (\theta_1,\ldots,\theta_p)^t$ and positive definite covariance matrix Σ. It is desired to estimate θ by an estimator $\delta = (\delta_1,\ldots,\delta_p)^t$ under the quadratic loss

$$L(\delta,\theta,\Sigma) = (\delta-\theta)^t Q(\delta-\theta)/tr(Q\Sigma),$$

where Q is a positive definite (p×p) matrix and "tr" stands for trace. As usual, an estimator will be evaluated in terms of its risk (expected loss)

$$R(\delta,\theta,\Sigma) = E_{\theta,\Sigma} L(\delta(X),\theta,\Sigma).$$

The usual minimax and best invariant estimator of θ is $\delta^0(X) = X$. Since Stein (1955) first showed that δ^0 could be improved upon for $Q = \Sigma = I$ (the identity matrix) a large body of work has developed concerned with finding significant improvements upon δ^0. For the most part these efforts have been directed towards problems where either Σ was known (or known up to a multiplicative constant) or where $Q = \Sigma^{-1}$ (a rather special situation). For the general situation, Berger and Bock (1976a) and (1977) and Shinozaki (1977) were able to find minimax estimators (better than δ^0) under the assumption that Σ was an unknown diagonal matrix or could be reduced to an unknown diagonal matrix. Berger, Bock, Brown, Casella, and Gleser (1977) (hereafter to be referred to as B^3CG (1977)) and Gleser (1979) succeeded in developing minimax estimators with no restrictions on Σ or Q. In this paper we broaden and strengthen the class of minimax estimators found earlier, and also provide a completely analytic proof of dominance.

It will be assumed that $\ddagger$ is completely unknown, but that a random $(p\times p)$ matrix W is available, from which $\ddagger$ can be estimated. The matrix W is assumed to have a Wishart distribution with parameter $\ddagger$ and n degrees of freedom (to be denoted $\mathscr{W}_p(\ddagger,n)$), and is assumed to be independent of X. It will also be assumed that $n > p+1$. We will consider estimators of θ having the form

$$\delta^c(X,W) = (I - c\alpha(Q^{\frac{1}{2}} WQ^{\frac{1}{2}})h(X^tW^{-1}X)Q^{-1}W^{-1})X, \tag{1.1}$$

where α and h are positive real functions satisfying certain properties and $c \geq 0$. These estimators will be shown to be minimax for $0 \leq c \leq c_{n,p}$, the constants $c_{n,p}$ being determined in the course of the investigation.

The particular case of (1.1) considered in B^3CG (1977) was

$$\delta^c(X,W) = (I - \frac{c\, ch_{min}(QW)}{(n-p-1)X^tW^{-1}X} Q^{-1}W^{-1})X, \tag{1.2}$$

corresponding to the choice $\alpha(Q^{\frac{1}{2}} WQ^{\frac{1}{2}}) = ch_{min}(Q^{\frac{1}{2}} WQ^{\frac{1}{2}})/(n-p-1)$ and $h(X^tW^{-1}X) = 1/(X^tW^{-1}X)$. (We use $ch\min(A)$ to denote the minimum characteristic root of the matrix A.) Gleser (1979) considered more general functions h, but also was restricted to the above choice of α.

The other, perhaps more useful, generalizations obtained here concern the constants $c_{n,p}$. First of all, larger values of $c_{n,p}$ are obtained (for $n-p \geq 12$) than in the B^3CG (1977) or Gleser (1979) papers. Hence larger choices of c in (1.1) or (1.2) are allowed, which can lead to substantial improvements. Secondly, a purely analytic result is obtained, namely that (1.1) is minimax if

$$0 \leq c \leq c^*_{n,p} = 2p - 4(n-p+1)\lambda_{n,p}, \tag{1.3}$$

where

$$\lambda_{n,p} = E[1/ch_{min}(V)], \tag{1.4}$$

V being a $\mathcal{W}_p(I,n)$ random matrix. Previous proofs of minimaxity obtained the values $c_{n,p}$ only as solutions to intractable functional equations involving Wishart matrices, the solutions of which were estimated by simulation. This was unsatisfactory to some statisticians, because the $c_{n,p}$ were not guaranteed by theory to be positive, and hence one must "trust" the simulation to believe that better minimax estimators were indeed found. Of course, the chance that the simulations gave drastically wrong values of $c_{n,p}$ is negligible, but it is worthwhile to have a purely analytic proof that if (1.3) is satisfied then δ^c is minimax. Unfortunately, the $c^*_{n,p}$ in (1.3) are considerably smaller than the $c_{n,p}$ found by the simulation technique, and hence give a much smaller class of minimax estimators.

The proofs of the minimax results in this paper are themselves of interest because they combine two methodologies that have been develped in simultaneous estimation problems: the integration by parts technique (also called the unbiased estimate of risk technique) developed by Stein (1973) to deal with estimating θ when Σ is known, and techniques developed independently by Haff (1977, 79) and Stein (unpublished) which lead to Wishart identities useful in the estimation of Σ. This combination of techniques should be of considerable use to other researchers in the area.

Throughout the paper, E will stand for expectation, with subscripts indicating parameter values at which the expectation is to be taken and superscripts indicating the random variables or distribution with respect to which the expectation is to be taken. When no confusion can arise, subscripts and superscripts of E may be omitted.

2. A Stochastic Ordering Result

In this section a result concerning the stochastic ordering of conditional distributions of Wishart characteristic roots is established. This result will be needed in the proof of the minimax theorem. After developing the result below, it was brought to our attention that the relevant stochastic ordering property was proven in Dykstra and Hewett (1978) by a different line of reasoning. We give our original result below because it deals with a somewhat broader class of distributions than just that considered in Dykstra and Hewett (1978), and hence may be of independent interest, and also because the purely analytic proof may be useful elsewhere.

It will be assumed that the random vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)$ has a joint density (with respect to Lebesgue measure) of the form

$$f(\lambda) = [\prod_{i=1}^{p} g_i(\lambda_i)][\prod_{(s,t)\in A} (\lambda_s - \lambda_t)] I_{\{\lambda : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p\}}(\lambda), \qquad (2.1)$$

where $g_i(\lambda_i) \geq 0$, $I_\Omega(\lambda)$ is the indicator function on Ω (i.e., $I_\Omega(\lambda) = 1$ if $\lambda \in \Omega$ and $I_\Omega(\lambda) = 0$ if $\lambda \notin \Omega$), and A is any set of pairs (s,t) such that s and t are integers between 1 and p, $s < t$, and $(s,s+1) \in A$ for $s = 1,\ldots,p-1$. The density of decreasingly ordered Wishart roots is in this class.

<u>Theorem 1.</u> If λ has a density of the form (2.1), then, for $j=1,\ldots,p-1$, the conditional distribution of $(\lambda_1,\ldots,\lambda_j)$ given $(\lambda_{j+1},\ldots,\lambda_p)$ is stochastically nondecreasing in $\lambda_k (k=j+1,\ldots,\lambda_p)$, in the sense that

$$E[\phi(\lambda_1,\ldots,\lambda_j)|\lambda_{j+1},\ldots,\lambda_p]$$

is nondecreasing in λ_k for any real valued function ϕ which is nondecreasing in the $\lambda_\ell (\ell=1,\ldots,p)$ (and for which the expectation exists).

<u>Proof.</u> From Proposition C.1 of Chapter 17 of Marshall and Olkin (1979) (a proposition based on results on stochastic ordering due to Veinott (1965) and Kamae, Krengel, and O'Brien (1977)) it suffices to show that, for $j=1,\ldots,p-1$ and all $t > \lambda_{j+1}$, $P(\lambda_j > t|\lambda_{j+1},\ldots,\lambda_p)$ is nondecreasing in λ_k,

$k=j+1,\ldots,p$. Defining $d\underset{\sim}{\lambda}^j = d\lambda_1 d\lambda_2 \cdots d\lambda_j$ and $A_j = \{(s,t)\in A: s\leq j\}$, it is clear that

$$P(\lambda_j>t|\lambda_{j+1},\ldots,\lambda_p)$$

$$= \frac{\int_t^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty f(\underset{\sim}{\lambda})d\underset{\sim}{\lambda}^j}{\int_{\lambda_{j+1}}^\infty \int_{\lambda_j}^\infty \int_{\lambda_2}^\infty f(\underset{\sim}{\lambda})d\underset{\sim}{\lambda}^j}$$

$$= \frac{\int_t^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty [\prod_{i=1}^j g_i(\lambda_i)] \prod_{(s,t)\in A_j} (\lambda_s-\lambda_t)d\underset{\sim}{\lambda}^j}{\int_{\lambda_{j+1}}^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty [\prod_{i=1}^j g_i(\lambda_i)] \prod_{(s,t)\in A_j} (\lambda_s-\lambda_t)d\underset{\sim}{\lambda}^j} . \quad (2.2)$$

To show that the last expression is nondecreasing in $\lambda_k(k=j+1,\ldots,p)$, we will simply show that its partial derivative with respect to λ_k is non-negative. The notation will be considerably simplified if we define

$$I_{j,k} = \{s: (s,t)\in A_j \text{ and } t=k\},$$

$$\varphi_{j,k}(\underset{\sim}{\lambda}) = [\prod_{i=1}^j g_i(\lambda_i)][\prod_{(s,t)\in A_j} (\lambda_s-\lambda_t)]/\prod_{s\in I_{j,k}} (\lambda_s-\lambda_k),$$

and, for $\ell\in I_{j,k}$,

$$\rho_{j,k}^\ell(\underset{\sim}{\lambda}) = [\prod_{s\in I_{j,k}} (\lambda_s-\lambda_k)]/(\lambda_\ell-\lambda_k).$$

Note that $\varphi_{j,k}(\underset{\sim}{\lambda})$ does not depend on λ_k, and that

$$\frac{\partial}{\partial\lambda_k} \prod_{s\in I_{j,k}} (\lambda_s-\lambda_k) = -\sum_{s\in I_{j,k}} \rho_{j,k}^s(\underset{\sim}{\lambda}).$$

Thus, letting N and D stand for the numerator and denominator, respectively, of the last expression in (2.2), it follows that

$$\frac{\partial}{\partial\lambda_k} P(\lambda_j > t \mid \lambda_{j+1},\dots,\lambda_p)$$

$$= D^{-1} \int_t^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty \varphi_{j,k}(\underset{\sim}{\lambda})[-\sum_{s\in I_{j,k}} \rho^s_{j,k}(\underset{\sim}{\lambda})]d\underset{\sim}{\lambda}^j$$

$$- ND^{-2} \int_{\lambda_{j+1}}^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty \varphi_{j,k}(\underset{\sim}{\lambda})[-\sum_{s\in I_{j,k}} \rho^s_{j,k}(\underset{\sim}{\lambda})]d\underset{\sim}{\lambda}^j.$$

(When k=j+1, the boundary term arising from differentiating $\int_{\lambda_{j+1}}^\infty$ is zero, since $(j,j+1)\in A_j$, and so the integrand of D evaluated at λ_{j+1} is zero.) Defining

$$h^s_{j,k}(\underset{\sim}{\lambda}) = \varphi_{j,k}(\underset{\sim}{\lambda})\rho^s_{j,k}(\underset{\sim}{\lambda}),$$

and observing that

$$[\prod_{i=1}^j g_i(\lambda_i)][\prod_{(s,t)\in A_j} (\lambda_s-\lambda_t)] = h^s_{j,k}(\underset{\sim}{\lambda})(\lambda_s-\lambda_k),$$

it is clear that

$$\frac{\partial}{\partial\lambda_k} P(\lambda_j > t \mid \lambda_{j+1},\dots,\lambda_p)$$

$$= D^{-2} \sum_{s\in I_{j,k}} \Big\{ [\int_t^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty h^s_{j,k}(\underset{\sim}{\lambda})(\lambda_s-\lambda_k)d\underset{\sim}{\lambda}^j][\int_{\lambda_{j+1}}^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty h^s_{j,k}(\underset{\sim}{\lambda})d\underset{\sim}{\lambda}^j]$$

$$- [\int_{\lambda_{j+1}}^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty h^s_{j,k}(\underset{\sim}{\lambda})(\lambda_s-\lambda_k)d\underset{\sim}{\lambda}^j][\int_t^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty h^s_{j,k}(\underset{\sim}{\lambda})d\underset{\sim}{\lambda}^j] \Big\}.$$

Breaking up the integrals $\int_{\lambda_{j+1}}^\infty$ into $\int_{\lambda_{j+1}}^t + \int_t^\infty$ in the above expression and subtracting common terms gives

$$\frac{\partial}{\partial\lambda_k} P(\lambda_j>t|\lambda_{j+1},\ldots,\lambda_p)$$

$$= D^{-2} \sum_{s\in I_{j,k}} \Big\{[\int_t^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty h^s_{j,k}(\underset{\sim}{\lambda})(\lambda_s-\lambda_k)d\underset{\sim}{\lambda}^j][\int_{\lambda_{j+1}}^t \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty h^s_{j,k}(\underset{\sim}{\lambda})d\underset{\sim}{\lambda}^j]$$

$$- [\int_{\lambda_{j+1}}^t \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty h^s_{j,k}(\underset{\sim}{\lambda})(\lambda_s-\lambda_k)d\underset{\sim}{\lambda}^j][\int_t^\infty \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty h^s_{j,k}(\underset{\sim}{\lambda})d\underset{\sim}{\lambda}^j]\Big\}. \quad (2.3)$$

Using (2.3), we can now complete the proof by induction on j. Thus consider first j=1, and observe that $I_{1,k}=\{1\}$ (since $s\le j$ if $s\in I_{j,k}$). Hence

$$\frac{\partial}{\partial\lambda_k} P(\lambda_1>t|\lambda_2,\ldots,\lambda_p) = D^{-2}\Big\{[\int_t^\infty h^1_{1,k}(\underset{\sim}{\lambda})(\lambda_1-\lambda_k)d\lambda_1][\int_{\lambda_2}^t h^1_{1,k}(\underset{\sim}{\lambda})d\lambda_1]$$

$$- [\int_{\lambda_2}^t h^1_{1,k}(\underset{\sim}{\lambda})(\lambda_1-\lambda_k)d\lambda_1][\int_t^\infty h^1_{1,k}(\underset{\sim}{\lambda})d\lambda_1]\Big\}$$

$$\ge D^{-2}\Big\{[\int_t^\infty h^1_{1,k}(\underset{\sim}{\lambda})(t-\lambda_k)d\lambda_1][\int_{\lambda_2}^t h^1_{1,k}(\underset{\sim}{\lambda})d\lambda_1]$$

$$- [\int_{\lambda_2}^t h^1_{1,k}(\underset{\sim}{\lambda})(t-\lambda_k)d\lambda_1][\int_t^\infty h^1_{1,k}(\underset{\sim}{\lambda})d\lambda_1]\Big\}$$

$$= 0 \quad \text{(factoring out } (t-\lambda_k)).$$

This establishes that $P(\lambda_j>t|\lambda_{j+1},\ldots,\lambda_p)$ is nondecreasing in λ_k for j=1.

Assume now the induction hypothesis, that

$$P(\lambda_\ell>t|\lambda_{\ell+1},\ldots,\lambda_p) \text{ is nondecreasing in } \lambda_k \ (k=\ell+1,\ldots,p)$$
$$\text{for } \ell=1,2,\ldots,j-1. \quad (2.4)$$

We must then show that this is also true for $\ell=j$. To this end, define

$$\psi^s_{j,k}(\lambda_j,\ldots,\lambda_p) = \int_{\lambda_j}^\infty \cdots \int_{\lambda_2}^\infty h^s_{j,k}(\underset{\sim}{\lambda})(\lambda_s-\lambda_k)d\underset{\sim}{\lambda}^{j-1},$$

and

$$\Delta_{j,k}^{s}(\lambda_j,\ldots,\lambda_p) = \int_{\lambda_j}^{\infty} \cdots \int_{\lambda_2}^{\infty} h_{j,k}^{s}(\underset{\sim}{\lambda})d\underset{\sim}{\lambda}^{j-1}.$$

Observe that

$$\frac{\psi_{j,k}^{s}(\lambda_j,\ldots,\lambda_p)}{\Delta_{j,k}^{s}(\lambda_j,\ldots,\lambda_p)} = E^*[\lambda_s-\lambda_k],$$

where "*" refers to expectation with respect to the "new" density

$$h_{j,k}^{s}(\underset{\sim}{\lambda})/\int_{\lambda_j}^{\infty} \cdots \int_{\lambda_2}^{\infty} h_{j,k}^{s}(\underset{\sim}{\lambda})d\underset{\sim}{\lambda}^{j-1} \tag{2.5}$$

on $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{j-1}$. For $s < j$, observe that the density in (2.5) is the same as the conditional density of $(\lambda_1,\ldots,\lambda_{j-1})$ given $(\lambda_j,\ldots,\lambda_p)$ when $\underset{\sim}{\lambda} = (\lambda_1,\ldots,\lambda_p)$ is assumed to have the density

$$f_{s,k}^{*}(\underset{\sim}{\lambda}) = cf(\underset{\sim}{\lambda})/(\lambda_s-\lambda_k),$$

c being the appropriate normalizing constant. Since $s < j$ and $k \geq j + 1$, $f_{s,k}^{*}$ is a density of the form (2.1). Hence by the induction hypothesis (2.4) (applied to $f_{s,k}^{*}$) together with another application of Marshall and Olkin (1979), it follows that the $f_{s,k}^{*}$ conditional density of $(\lambda_1,\ldots,\lambda_{j-1})$ given $(\lambda_j,\ldots,\lambda_p)$ is stochastically nondecreasing. Since $[\lambda_s-\lambda_k]$ is non-decreasing in λ_s ($s < j < k$ recall), it can be concluded that $E^*[\lambda_s-\lambda_k]$ is nondecreasing in λ_j. When $s=j$, $E^*[\lambda_j-\lambda_k] = \lambda_j - \lambda_k$, which is trivially nondecreasing in λ_j. Hence in all cases it can be concluded that

$$\frac{\psi_{j,k}^{s}(\lambda_j,\ldots,\lambda_p)}{\Delta_{j,k}^{s}(\lambda_j,\ldots,\lambda_p)} \text{ is nondecreasing in } \lambda_j. \tag{2.6}$$

To complete the argument, a use of (2.6) in (2.3) gives (simply writing ψ^s and Δ^s when convenient)

$$\frac{\partial}{\partial\lambda_k} P(\lambda_j>t|\lambda_{j+1},\ldots,\lambda_p)$$

$$= D^{-2} \sum_{s\in I_{j,k}} \left\{[\int_t^\infty \Delta^s(\frac{\psi^s}{\Delta^s})d\lambda_j][\int_{\lambda_{j+1}}^t \Delta^s d\lambda_j]-[\int_{\lambda_{j+1}}^t \Delta^s(\frac{\psi^s}{\Delta^s})d\lambda_j][\int_t^\infty \Delta^s d\lambda_j]\right\}$$

$$\geq D^{-2} \sum_{s\in I_{j,k}} [\int_t^\infty \Delta^s(\lambda_j,\ldots,\lambda_p)(\frac{\psi^s(t,\lambda_{j+1},\ldots,\lambda_p)}{\Delta^s(t,\lambda_{j+1},\ldots,\lambda_p)})d\lambda_j][\int_{\lambda_{j+1}}^t \Delta^s d\lambda_j]$$

$$-[\int_{\lambda_{j+1}}^t \Delta^s(\lambda_j,\ldots,\lambda_p)(\frac{\psi^s(t,\lambda_{j+1},\ldots,\lambda_p)}{\Delta^s(t,\lambda_{j+1},\ldots,\lambda_p)})d\lambda_j][\int_t^\infty \Delta^s d\lambda_j]$$

$$= 0 \text{ (factoring out the constant term } \frac{\psi^s(t,\lambda_{j+1},\ldots,\lambda_p)}{\Delta^s(t,\lambda_{j+1},\ldots,\lambda_p)} \text{)}.$$

This completes the induction step (verifying (2.4) for $\ell=j$) and hence the proof. ||

It is interesting to note that the induction step in the above proof necessitated working with a new density of the form (2.1), and hence would not have been possible had we tried to consider only the Wishart situation.

The particular results that will be needed in the minimax theorem are stated below as corollaries.

<u>Corollary 1.</u> If $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the characteristic roots of a $\mathscr{W}_p(I,n)$ matrix, and $\phi(\lambda_1,\ldots,\lambda_p)$ is a real valued function which is nondecreasing in each coordinate, then for $j=1,\ldots,p-1$

$$E[\phi(\lambda_1,\ldots,\lambda_p)|\lambda_{j+1},\ldots,\lambda_p]$$

is nondecreasing in λ_k, $k \geq j+1$.

<u>Proof.</u> Let $\lambda_p'' > \lambda_p'$. The density of $(\lambda_1,\ldots,\lambda_p)$ is of the form (2.1), so, by Theorem 1,

$$E[\phi(\lambda_1,\ldots,\lambda_p')|\lambda_{j+1},\ldots,\lambda_p'] \le E[\phi(\lambda_1,\ldots,\lambda_p')|\lambda_{j+1},\ldots,\lambda_p''].$$

But $\phi(\lambda_1,\ldots,\lambda_p') \le \phi(\lambda_1,\ldots,\lambda_p'')$, so that

$$E[\phi(\lambda_1,\ldots,\lambda_p')|\lambda_{j+1},\ldots,\lambda_p'] \le E[\phi(\lambda_1,\ldots,\lambda_p')|\lambda_{j+1},\ldots,\lambda_p''] \le E[\phi(\lambda_1,\ldots,\lambda_p'')|\lambda_{j+1},\ldots,\lambda_p''],$$

establishing the result for k=p. The proof for other k is identical. ||

Corollary 2. If $\lambda_1 \ge \cdots \ge \lambda_p$ are the characteristic roots of a $\mathcal{W}_p(I,n)$ matrix, and $\phi(\lambda_1,\ldots,\lambda_p)$ and $\psi(\lambda_1,\ldots,\lambda_p)$ are real valued functions which are nondecreasing in each coordinate, then

$$E[\phi(\lambda_1,\ldots,\lambda_p)\psi(\lambda_1,\ldots,\lambda_p)] \ge E[\phi(\lambda_1,\ldots,\lambda_p)]E[\psi(\lambda_1,\ldots,\lambda_p)].$$

Proof. Write

$$E[\phi\psi] = E^{\lambda_2,\ldots,\lambda_p}E^{\lambda_1|\lambda_2,\ldots,\lambda_p}[\phi\psi],$$

where the first expectation on the right is with respect to the marginal distribution of $(\lambda_2,\ldots,\lambda_p)$ and the second with respect to the conditional distribution of λ_1 given $\lambda_2,\ldots,\lambda_p$. Since ϕ and ψ are nondecreasing in λ_1,

$$E^{\lambda_1|\lambda_2,\ldots,\lambda_p}[\phi\psi] \ge E^{\lambda_1|\lambda_2,\ldots,\lambda_p}[\phi]E^{\lambda_1|\lambda_2,\ldots,\lambda_p}[\psi].$$

Define the quantities on the right hand side above as $\phi^{(1)}(\lambda_2,\ldots,\lambda_p)$ and $\psi^{(1)}(\lambda_2,\ldots,\lambda_p)$, respectively. Corollary 1 (used with j=1) shows that $\phi^{(1)}$ and $\psi^{(1)}$ are nondecreasing in their coordinates. Hence

$$E^{\lambda_2,\ldots,\lambda_p}[\phi^{(1)}\psi^{(1)}] = E^{\lambda_3,\ldots,\lambda_p}E^{\lambda_2|\lambda_3,\ldots,\lambda_p}[\phi^{(1)}\psi^{(1)}]$$

$$\ge E^{\lambda_3,\ldots,\lambda_p}\left\{E^{\lambda_2|\lambda_3,\ldots,\lambda_p}[\phi^{(1)}]E^{\lambda_2|\lambda_3,\ldots,\lambda_p}[\psi^{(1)}]\right\}.$$

Define the quantities inside the curly brackets above as $\phi^{(2)}(\lambda_3,\ldots,\lambda_p)$ and

$\psi^{(2)}(\lambda_3,\ldots,\lambda_p)$. Since

$$\phi^{(2)}(\lambda_3,\ldots,\lambda_p) = E^{\lambda_1,\lambda_2|\lambda_3,\ldots,\lambda_p}[\phi^{(1)}],$$

Corollary 1 can again be applied to conclude that $\phi^{(2)}$ is nondecreasing in its coordinates. The same conclusion holds for $\psi^{(2)}$. Continuing in the obvious way and combining the inequalities gives the desired result.||

3. Minimaxity of δ^c

In this section, the estimator defined by (1.1) is shown to be minimax under certain conditions. The needed conditions on the functions h and α are as follows.

Condition h. Assume the function h: $[0,\infty) \to [0,\infty)$ satisfies

(i) $0 \leq zh(z) \leq 1$;

(ii) h is continuous and piecewise differentiable;

(iii) $-zh'(z)/h(z) \leq 1$, where $h'(z) = (d/dz)h(z)$;

(iv) $h(z)$ is nonincreasing.

Condition α. Assume the real valued function $\alpha(S)$, defined on the cone of positive definite $(p\times p)$ matrices S (whose elements are s_{ij}), satisfies

(i) $0 \leq \alpha(S) \leq [ch_{min}(S)]/(n-p-1)$;

(ii) $\alpha(S)$ is nondecreasing in the characteristic roots of S;

(iii) $\alpha(S)$ is continuous and differentiable;

(iv) $tr\{[D \log \alpha(S)]\ddagger\} \leq ch_{max}(\ddagger S^{-1})$, where "$ch_{max}$" denotes maximum characteristic root and D is the matrix differential operator wherein the (i,j) element of Dg(S) is defined to be

$$[Dg(S)]_{i,j} = \begin{cases} \frac{\partial}{\partial s_{ij}} g(S) & \text{if} \quad i = j \\ \frac{\partial}{\partial s_{ij}} [\tfrac{1}{2} g(S)] & \text{if} \quad i \neq j \end{cases}$$

Some natural and useful choices of α and h which satisfy the above conditions are given in the following lemmas.

Lemma 1. The function $h(z) = \min(k, \frac{1}{z})$ (where $k \geq 0$) satisfies Condition h.

Proof. Easy calculation. ||

Observe that the use of h of the above form in (1.1) eliminates a glaring deficiency of the estimator (1.2), namely that (1.2) blows up (i.e., $|\delta^C| \to \infty$) as $|X| \to 0$.

Lemma 2. The following choices of α satisfy Condition α:

(a) $\alpha(S) = [ch_{min}(S)]/(n-p-1)$;

(b) $\alpha(S) = 1/[(n-p-1)tr(S^{-1})]$.

Proof. Only Condition α(iv) is not immediately verifiable.

(a) From Wilkinson (1965), it can be seen that

$$D\, ch_{min}(S) = R_p R_p^t,$$

where R_p is the normalized eigenvector corresponding to $ch_{min}(S)$ (or, equivalently, the last column of the orthogonal matrix R for which $S = R\Lambda R^t$, Λ being the diagonal matrix of decreasingly ordered characteristic roots of S). Hence

$$\begin{aligned} tr\{[D \log \alpha(S)]\ddagger\} &= \frac{1}{\alpha(S)} tr\{[D\alpha(S)]\ddagger\} \\ &= \frac{1}{ch_{min}(S)} tr\{R_p R_p^t \ddagger\} \\ &= \frac{1}{ch_{min}(S)} R_p^t \ddagger R_p \end{aligned}$$

$$= \frac{R_p^t \Sigma R_p}{R_p^t S R_p}$$

$$\leq ch_{max}(\Sigma S^{-1}),$$

completing the verification of Condition α(iv).

(b) An application of Lemma 4 of Haff (1980b) shows that $[D\,tr\,S^{-1}] = -S^{-2}$. Hence

$$tr\{[D \log \alpha(S)]\Sigma\} = (tr\,S^{-1})\,tr\{[D(\frac{1}{tr\,S^{-1}})]\Sigma\}$$

$$= (tr\,S^{-1})(\frac{-1}{(tr\,S^{-1})^2})\,tr\{[D\,tr\,S^{-1}]\Sigma\}$$

$$= \frac{-1}{tr\,S^{-1}} \cdot tr\{-S^{-2}\Sigma\}$$

$$= \frac{tr\{S^{-\frac{1}{2}} B S^{-\frac{1}{2}}\}}{tr\,S^{-1}} \cdot ch_{max}(\Sigma S^{-1}),$$

where $B = S^{-\frac{1}{2}} \Sigma S^{-\frac{1}{2}}/ch_{max}(\Sigma S^{-1})$. Since B is positive definite and has roots less than or equal to one, it follows that the roots of $S^{-\frac{1}{2}} B S^{-\frac{1}{2}}$ are less than or equal to the roots of S^{-1}. Hence

$$\frac{tr\{S^{-\frac{1}{2}} B S^{-\frac{1}{2}}\}}{tr\,S^{-1}} \leq 1,$$

completing the verification of Condition α(iv). ||

The following theorem contains the heart of the minimax analysis.

<u>Theorem 2.</u> The estimator $\delta^c(X,W)$ in (1.1) is minimax (and hence $R(\delta^c,\theta,\Sigma) \leq R(\delta^0,\theta,\Sigma)$ for all θ and Σ) if

$$0 \leq E^V[\alpha(AVA)h(\rho v_1)\{2p-4\lambda_{n,p}-c-4(n-p)v_1(1+\ell^t V_2^{-1}\ell)\}] \qquad (3.1)$$

for all $\rho \geq 0$ and positive definite matrices A, where $\lambda_{n,p}$ is defined in (1.4), $v_1^{-1} \sim \chi^2(n-p+1)$, $\ell \sim \mathcal{N}_{p-1}(0,I)$, $V_2 \sim \mathcal{W}_{p-1}(n,I)$, and $V \sim \mathcal{W}_p(I,n)$

is defined by

$$V = \begin{pmatrix} v_1^{-1} + |\ell|^2 & \ell^t V_2^{\frac{1}{2}} \\ V_2^{\frac{1}{2}}\ell & V_2 \end{pmatrix} . \tag{3.2}$$

(The above decomposition of V is well known, and v_1, ℓ, and V_2 are independent.)

Proof. We want to show that

$$0 \geq \Delta_c \equiv \Delta_c(\theta,\ddagger) \equiv R(\delta^c,\theta,\ddagger) - R(\delta^0,\theta,\ddagger)$$

$$= E_{\theta,\ddagger}^{X,W}[\{\delta^c(X,W)-\theta\}^t Q\{\delta^c(X,W)-\theta\}-\{\delta^0(X)-\theta\}^t Q\{\delta^0(X)-\theta\}].$$

Observe first that, without loss of generality, we can assume that the matrix Q in the quadratic loss is the identity. This can be seen by considering the transformations $X^* = Q^{\frac{1}{2}} X$, $\theta^* = Q^{\frac{1}{2}}\theta$, $\delta^{c*} = Q^{\frac{1}{2}}\delta^c$, $\ddagger^* = Q^{\frac{1}{2}} \ddagger Q^{\frac{1}{2}}$, and $W^* = Q^{\frac{1}{2}} W Q^{\frac{1}{2}}$. It is easy to check that risks in the * problem are simply reparametrizations of risks in the original problem, and that the * problem has exactly the same structure as the original problem except that the loss in the * problem is

$$(\delta^*-\theta^*)^t(\delta^*-\theta^*)/tr(\ddagger^*).$$

We will henceforth assume that $Q = I$.

Expanding the quadratic loss in the above expression for Δ_c verifies that

$$\Delta_c = -E_{\theta,\ddagger}^{X,W}\{2c\alpha(W)h(X^tW^{-1}X)(X-\theta)^tW^{-1}X-c^2\alpha^2(W)h^2(X^tW^{-1}X)X^tW^{-2}X\}. \tag{3.3}$$

As in Berger (1976), integrating by parts with respect to the X_i $(i=1,\ldots,p)$ shows that

$$E[h(X^tW^{-1}X)(X-\theta)^tW^{-1}X] = E[h(X^tW^{-1}X)tr(W^{-1}\ddagger)+2h'(X^tW^{-1}X)X^tW^{-1}\ddagger W^{-1}X].$$

Condition h can easily be shown to ensure the validity of the integration by parts for $p \geq 3$. (The usefulness of such an integration by parts was first noticed by Stein (1973).) Using this in (3.3) gives

$$\Delta_c = -E\{2c\alpha(W)h(X^tW^{-1}X)tr(W^{-1}\Sigma)$$

$$+ 4c\alpha(W)h'(X^tW^{-1}X)X^tW^{-1}\Sigma W^{-1}X - c^2\alpha^2h^2X^tW^{-2}X\} . \quad (3.4)$$

At this point, we need the following identity, due to Haff (1977) and Stein (unpublished work):

$$E_{\Sigma}^{W}[trW^{-1}B] = \frac{1}{n-p-1}\{E[tr(\Sigma^{-1}B)]-2E[tr(DB)]\}, \quad (3.5)$$

where B is a (p×p) matrix function of W and Σ, and DB is defined as the matrix formally obtained by multiplying the (p×p) matrix D (whose elements will be denoted d_{ij}) times the matrix B, with the interpretation that $d_{ij}b$ (b an element of B) in the formal product stands for

$$d_{ij}b = \begin{cases} \frac{\partial}{\partial s_{ij}} b & \text{if} \quad i = j \\ \frac{\partial}{\partial s_{ij}} (\tfrac{1}{2} b) & \text{if} \quad i \neq j . \end{cases}$$

(The definition of D here is equivalent to that in Condition α, providing we recognize $Dg(S)$ as shorthand notation for $D(g(S)I)$.) Choosing $B = \alpha(W)h(X^tW^{-1}X)\Sigma$, it follows that

$$E[\alpha(W)h(X^tW^{-1}X)tr(W^{-1}\Sigma)]$$

$$= \frac{1}{n-p-1}\left\{pE[\alpha(W)h(X^tW^{-1}X)]-2E[tr(D\{\alpha h\Sigma\})]\right\} . \quad (3.6)$$

(It can be shown that this choice of B satisfies the conditions for (3.5) to be valid, when Condition α and Condition h hold.) Observe that

$$D\{\alpha(W)h(X^tW^{-1}X)\Sigma\} = [D\{\alpha(W)h(X^tW^{-1}X)\}]\Sigma$$

$$= \alpha(W)[Dh(X^tW^{-1}X)]\Sigma + h(X^tW^{-1}X)[D\alpha(W)]\Sigma$$

$$= \alpha(W)h'(X^tW^{-1}X)[DX^tW^{-1}X]\Sigma + h[D\alpha]\Sigma$$

$$= \alpha(W)h'(X^tW^{-1}X)[D\{tr(XX^tW^{-1})\}]\ddagger + h[D\alpha]\ddagger$$

$$= -\alpha(W)h'(X^tW^{-1}X)W^{-1}XX^tW^{-1}\ddagger + h[D\alpha]\ddagger \ ,$$

the last step following from Lemma 4 of Haff (1980b). Using this in (3.6), applying (3.6) to the first term in (3.4), and noting that

$$\frac{1}{\alpha(W)}[D\alpha(W)] = D\log\alpha(W),$$

establishes the equality

$$\Delta_C = -E[c\alpha(W)h(X^tW^{-1}X)\{\frac{2p}{n-p-1} + 4(1+\frac{1}{n-p-1})\frac{h'(X^tW^{-1}X)}{h(X^tW^{-1}X)}(X^tW^{-1}\ddagger W^{-1}X)$$

$$-\frac{4}{n-p-1}tr([D\log\alpha]\ddagger)-c\alpha(W)h(X^tW^{-1}X)(X^tW^{-2}X)\}]. \qquad (3.7)$$

Condition α(i) and Condition h(i) imply that

$$\alpha(W)h(X^tW^{-1}X)(X^tW^{-2}X) \le \frac{ch_{min}(W)}{(n-p-1)} \cdot \frac{X^tW^{-2}X}{X^tW^{-1}X}$$

$$\le \frac{1}{n-p-1}ch_{min}(W)ch_{max}(W^{-1})$$

$$= \frac{1}{n-p-1} \ . \qquad (3.8)$$

Condition α(iv) gives that

$$tr\{[D\log\alpha(W)]\ddagger\} < ch_{max}(\ddagger W^{-1}), \qquad (3.9)$$

and Condition h(iii) implies that

$$-\frac{h'(X^tW^{-1}X)(X^tW^{-1}X)}{h(X^tW^{-1}X)} \cdot \frac{(X^tW^{-1}\ddagger W^{-1}X)}{(X^tW^{-1}X)} \le \frac{X^tW^{-1}\ddagger W^{-1}X}{X^tW^{-1}X} \ . \qquad (3.10)$$

Using (3.8), (3.9), and (3.10) in (3.7), together with the conditions that α, h, and c are nonnegative, results in the inequality

$$\Delta_C \le -E^{X,W}_{0,\ddagger}[c\alpha(W)h(X^tW^{-1}X)\{\frac{2p-c}{(n-p-1)} - \frac{4}{(n-p-1)}ch_{max}(\ddagger W^{-1})$$

$$-4(1+\frac{1}{n-p-1})\frac{X^tW^{-1}\ddagger W^{-1}X}{X^tW^{-1}X}\}] \ .$$

Making the transformations $Y = \Sigma^{-\frac{1}{2}} X$ and $S = \Sigma^{-\frac{1}{2}} W \Sigma^{-\frac{1}{2}} \sim \mathcal{W}_p(I,n)$ above, and simplifying, results in the equivalent expression

$$\Delta_c \leq - \frac{c}{(n-p-1)} E^{Y,S}[\alpha(\Sigma^{\frac{1}{2}} S \Sigma^{\frac{1}{2}}) h(Y^t S^{-1} Y)\{2p-c-4ch_{max}(S^{-1}) - 4(n-p) \frac{Y^t S^{-2} Y}{Y^t S^{-1} Y}\}]. \qquad (3.11)$$

Next, let Γ_Y be the orthogonal matrix such that $\Gamma_Y Y = (|Y|,0,\ldots,0)^t$, define $V = \Gamma_Y S \Gamma_Y^t \sim \mathcal{W}_p(I,n)$ (conditional on Y), and decompose V as in (3.2). It follows that (3.11) can be written

$$\Delta_c \leq - \frac{c}{(n-p-1)} E^Y E^V[\alpha(\Sigma^{*\frac{1}{2}} V \Sigma^{*\frac{1}{2}}) h(|Y|^2 v_1) \times \{2p-c-4ch_{max} V^{-1} - 4(n-p)(V^{-2})_{11}/v_1\}] ,$$

where $\Sigma^* = \Gamma_Y \Sigma \Gamma_Y^t$. Note that we have used the independence of Y and S, and will henceforth be working only with the second expectation, treating Y as fixed. Using the fact that $(V^{-2})_{11} = v_1^2(1+\ell^t V_2^{-1}\ell)$, it follows that $\Delta_c \leq 0$ if, for all $A = \Sigma^{*\frac{1}{2}}$ and $\rho = |Y|^2$,

$$E^V[\alpha(AVA)h(\rho v_1)\{2p-c-4ch_{max}(V^{-1})-4(n-p)v_1(1+\ell^t V_2^{-1}\ell)\}] \geq 0 . \qquad (3.12)$$

At this point, the following result is needed.

<u>Lemma 3.</u> If A is positive definite and $\rho \geq 0$, then

$$E^V[\alpha(AVA)h(\rho v_1)ch_{max}(V^{-1})] \leq E^V[\alpha(AVA)h(\rho v_1)]E^V[ch_{max}(V^{-1})] . \qquad (3.13)$$

To prove this, write $V = R\Lambda R^t$, where Λ is the diagonal matrix of eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ of V and R is an orthogonal $(p \times p)$ matrix. Note that the distributions of R and V are independent. Also, define $B = AR$, so that

$$E^V[\alpha(AVA)h(\rho v_1)ch_{max}(V^{-1})] = E^R E^\Lambda[\alpha(B\Lambda B^t)h(\rho v_1)\lambda_p^{-1}] . \qquad (3.14)$$

It is easy to see that the roots of $B\Lambda B^t$ are nondecreasing in the λ_i, so, by Condition α(ii), $\alpha(B\Lambda B^t)$ is nondecreasing in the λ_i. Also, v_1 can be written (letting $e_1 = (1,0,\ldots,0)^t$ and $Z = R^t e_1$)

$$v_1 = (V^{-1})_{11} = e_1^t V^{-1} e_1 = e_1^t R\Lambda^{-1} R^t e_1 = Z^t \Lambda^{-1} Z = \sum_{i=1}^{p} \frac{Z_i^2}{\lambda_i}, \qquad (3.15)$$

from which it is clear that v_1 is nonincreasing in the λ_i. Condition h(iv) then implies that $h(\rho v_1)$ is nondecreasing in the λ_i. Setting $\phi(\lambda_1,\ldots,\lambda_p) = \alpha(B\Lambda B^t)h(\rho v_1)$ and $\psi(\lambda_1,\ldots,\lambda_p) = -\lambda_p^{-1}$, it follows from Corollary 2 (Section 2) that

$$E^\Lambda[\phi\psi] \geq E^\Lambda[\phi]E^\Lambda[\psi].$$

Together with (3.14), this establishes the lemma.

The conclusion of the theorem follows by combining (3.12) and (3.13). ||

Corollary 3. The estimator $\delta^c(X,W)$ given in (1.1) is minimax providing (1.3) is satisfied.

Proof. For the situation of Theorem 1, observe that (letting $e_1 = (1,0,\ldots,0)^t$)

$$v_1(1+\ell^t V_{22}^{-1}\ell) = \frac{(V^{-2})_{11}}{(V^{-1})_{11}} = \frac{e_1^t V^{-2} e_1}{e_1^t V^{-1} e_1} \leq ch_{max}(V^{-1}).$$

Using this bound in (3.12), and then applying Lemma 3 as in Theorem 2 yields that the sufficient condition for minimaxity is

$$E^V[\alpha(\Lambda V\Lambda)h(\rho v_1)\{2p-c-4(n-p+1)\lambda_{n,p}\}] \geq 0.$$

This is clearly implied by (1.3). ||

Corollary 4. In the situation of Theorem 2, $\delta^c(X,W)$ is minimax if $0 \le c \le (2p-4\lambda_{n,p})$ and, for all $\rho \ge 0$ and positive semidefinite matrices A, either

(a.) $\alpha(S) = ch_{min}(S)/(n-p-1)$ and

$$0 \le E^V[\alpha(AVA)\{2p-4\lambda_{n,p}-c-\frac{4(n-p)}{(n-p-1)}(1+\ell^t V_2^{-1}\ell)\}]; \tag{3.16}$$

or

(b.) $h(z) = 1/z$ and

$$0 \le E^V[\alpha(AVA)\{2p-4\lambda_{n,p}-c-\frac{4(n-p)}{(n-p+1)}(1+\ell^t V_2^{-1}\ell)\}]. \tag{3.17}$$

Proof. Using (3.2), it is straightforward to verify that the roots of AVA are nonincreasing in v_1, and hence that $\alpha(AVA)$ is nonincreasing in v_1 (by Condition α(ii)). It follows that

$$E^{v_1}[\alpha(AVA)\{-v_1\}] \ge -E^{v_1}[\alpha(AVA)]E^{v_1}[v_1] = -(n-p-1)^{-1}E^{v_1}[\alpha(AVA)] \tag{3.18}$$

and

$$E^{v_1}[\alpha(AVA)v_1^{-1}] \ge E^{v_1}[\alpha(AVA)]E^{v_1}[v_1^{-1}] = (n-p+1)E^{v_1}[\alpha(AVA)]. \tag{3.19}$$

To verify the corollary for the situation of part (b), set $h(z) = 1/z$ in (3.1), and use (3.19), the assumption that $(2p-4\lambda_{n,p}-c) \ge 0$, and the independence of v_1, ℓ, and V_2 to conclude that

$$E^V[\alpha(AVA)(\rho v_1)^{-1}\{2p-4\lambda_{n,p}-c-4(n-p)v_1(1+\ell^t V_2^{-1}\ell)\}]$$
$$\ge \rho^{-1}E^V[\alpha(AVA)\{(2p-4\lambda_{n,p}-c)(n-p+1)-4(n-p)(1+\ell^t V_2^{-1}\ell)\}].$$

Assumption (3.17) ensures that this is positive, so (3.1) is satisfied.

To verify the corollary for the situation of part (a), observe from the relationship

$$V^{-1} = \begin{pmatrix} v_1 & -v_1 \ell^t V_2^{-\frac{1}{2}} \\ -v_1 V_2^{-\frac{1}{2}} \ell & V_2^{-1} + v_1 V_2^{-\frac{1}{2}} \ell \ell^t V_2^{-1} \end{pmatrix}$$

that $v_1^{-1}V^{-1}$ has roots which are nonincreasing in v_1. Hence $v_1 V$ has roots which are nondecreasing in v_1. It follows that

$$-v_1\alpha(AVA) = -ch_{\min}(v_1 AVA)/(n-p-1)$$

is nonincreasing in v_1. Since $\alpha(AVA)$ is nonincreasing in v_1 and $(2p-4\lambda_{n,p}-c) \geq 0$, it can be concluded that

$$\alpha(AVA)\{2p-4\lambda_{n,p}-c-4(n-p)v_1(1+\ell^t V_2^{-1}\ell)\}$$

is nonincreasing in v_1. Finally, $h(\rho v_1)$ is nonincreasing in v_1, so that

$$E^V[\alpha(AVA)h(\rho v_1)\{2p-4\lambda_{n,p}-c-4(n-p)v_1(1+\ell^t V_2^{-1}\ell)\}]$$

$$\geq E^{\ell,V_2}\left\{E^{v_1}[h(\rho v_1)]E^{v_1}[\alpha(AVA)\{2p-4\lambda_{n,p}-c-4(n-p)v_1(1+\ell^t V_2^{-1}\ell)\}]\right\}$$

$$\geq E^{v_1}[h(\rho v_1)]E^V[\alpha(AVA)\{2p-4\lambda_{n,p}-c-\frac{4(n-p)}{(n-p-1)}(1+\ell^t V_2^{-1}\ell)\}],$$

the last step following from (3.18). This last expression is positive by assumption (3.16), so (3.1) is satisfied and the proof is complete. ||

<u>Corollary 5.</u> Consider the estimator $\delta^c(X,W)$ in (1.1) with $\alpha(S) = ch_{\min}(S)/(n-p-1)$. This estimator is minimax if $0 \leq c \leq c_{n,p}$, where $c_{n,p}$ is the unique solution to

$$c = \min\left\{\frac{\tau_0(c)+\tau_1(c)}{\tau_0'(c)+\tau_1'(c)}, \frac{\tau_0(c)}{\tau_1(c)}\right\}, \tag{3.20}$$

where

$$\tau_0(c) = E^V[\rho(V)\{V_{22}I_{\Omega_c}(V)+ch_{\min}(V)I_{\bar{\Omega}_c}(V)\}] ,$$

$$\tau_1(c) = E^V[\rho(V)(V_{11}-V_{22})I_{\Omega_c}(V)] \ ,$$

$$\tau_0'(c) = E^V[V_{22}I_{\Omega_c}(V)+ch_{min}(V)I_{\bar{\Omega}_c}(V)] \ ,$$

$$\tau_1'(c) = E^V[(V_{11}-V_{22})I_{\Omega_c}(V)] \ ,$$

V_{ii} is the (i,i) element of $V \sim \mathscr{W}_p(I,n)$,

$$I_{\Omega_c}(V) = 1 - I_{\bar{\Omega}_c}(V) = \begin{cases} 1 & \text{if} \quad V \in \Omega_c \\ 0 & \text{if} \quad V \notin \Omega_c \end{cases} \ ,$$

$$\Omega_c = \{V: \rho(V) < c\} \ ,$$

and

(a.) $\rho(V) = 2p-4\lambda_{n,p} - \frac{4(n-p)}{(n-p-1)}(1+\ell^t V_2^{-1}\ell)$

<u>or</u>

(b.) $\rho(V) = 2p-4\lambda_{n,p} - \frac{4(n-p)}{(n-p+1)}(1+\ell^t V_2^{-1}\ell)$

when $h(z) = 1/z$.

<u>Proof.</u> The proof starts with (3.16) or (3.17) and then proceeds exactly as in B^3CG(1977). ||

Using Corollary 5, a simulation was performed to calculate the values $c_{n,p}$. It turned out that values of $c_{n,p}$, larger than those found in the B^3CG(1977) and Gleser (1979) papers, were obtained for $n - p \geq 12$ (roughly). (The analysis in the B^3CG(1977) and Gleser (1979) papers was, of course, different.) Table 1 gives the relevant constants $c_{n,p}$ for various n and p. The entries under "B" are the values of $c_{n,p}$ for the situation of Corollary 5(b) (which corresponds to the estimator (1.2) considered in the B^3CG(1977) paper), while the entries under "G" are the values of $c_{n,p}$ for the more general situation considered in Corollary 5(a) (which corresponds to the Gleser (1979) situation). The entries in Table 1 are the maximum of the values found using Corollary 5 and the values given in the earlier papers. The starred entries are those entries found using Corollary 5 which are bigger than the corresponding values found in B^3CG(1977) or Gleser (1979). (For $n < 16$ the B^3CG(1977) and Gleser (1979) values of $c_{n,p}$ were always larger than ours, so refer to those papers to find the appropriate $c_{n,p}$ for such n.) The improvement obtained using Corollary 5 was substantial, particularly for the Gleser (1979) situation. For $n = 30$ and $p = 17$, for instance, Corollary 5 gave $c_{n,p} = 19.00$, while the corresponding value in Gleser (1979) was 14.23. (The standard errors of the simulated solutions $c_{n,p}$ ranged from about .02 (for $p = 3$) to about .1 (for $n - p = 4$).)

TABLE 1

Values of $c_{n,p}$.

	n									
	16		18		20		25		30	
p	B	G	B	G	B	G	B	G	B	G
3	1.06*	.58	1.20*	.75	1.34*	.87	1.51	1.20	1.59*	1.27*
4	2.48*	1.79	2.65*	2.02	2.85*	2.25	3.09*	2.78	3.33*	2.98*
5	3.80	2.78	4.05*	3.17*	4.35*	3.63*	4.79*	4.28*	5.09*	4.71*
6	4.81	3.47	5.33	4.17*	5.73*	4.87*	6.43*	5.85*	6.80*	6.38*
7	5.78	3.93	6.42	5.05*	6.99*	5.96*	7.93*	7.29*	8.47*	8.00*
8	6.57	4.19	7.64	5.12	8.19	6.81*	9.26*	8.58*	10.15*	9.60*
9	7.02	3.86	8.40	5.56	9.22	7.38*	10.60	9.76*	11.80*	11.28*
10	6.79	3.66	8.90	5.17	10.25	6.80	11.98	10.92*	13.37*	12.62*
11	5.78	1.28	9.15	5.18	10.84	7.10	13.14	11.85*	14.74*	14.10*
12	2.73		8.42	4.21	11.10	6.52	14.20	12.66*	16.06*	15.35*
13			7.11	.94	11.09	6.25	15.48	12.66*	17.36*	16.40*
14			2.43		9.70	4.58	15.74	12.41*	18.72*	17.53*
15					7.93		16.61	11.77*	19.90*	18.57*
16					2.26		16.67	10.91	20.62	18.78*
17							16.67	10.45	21.56	19.00*
18							16.34	9.30	22.38	18.11*
19									22.83	14.84
20									23.47	14.52

As mentioned in the introduction, the more satisfying explicit bounds $c^*_{n,p}$ given by (1.3) (see also Corollary 3) are substantially less than the bounds in Table 1. This is indicated in Table 2, which gives values of $\lambda_{n,p} = E[1/ch_{min}(V)]$ (under the "λ" columns) and $c^*_{n,p}$ (under the "c" columns) for various n and p. (Only positive entries of $c^*_{n,p}$ are given. The bound was always negative for $n < 18$.) Of course, Corollary 3 does cover a broader class of estimators than Corollary 5.

TABLE 2

Values of $\lambda_{n,p}$ and $c^*_{n,p}$.

	n = 16		n = 18		n = 20		n = 25		n = 30	
p	λ	c	λ	c	λ	c	λ	c	λ	c
3	.129	-	.107	-	.092	-	.070	-	.054	-
4	.171	-	.136	-	.114	.25	.084	.61	.063	1.20
5	.221	-	.172	.37	.141	.98	.100	1.60	.073	2.41
6	.291	-	.215	.82	.171	1.74	.118	2.56	.084	3.60
7	.38	-	.27	1.04	.21	2.18	.138	3.51	.097	4.69
8	.50	-	.35	.60	.26	2.43	.163	4.26	.110	5.88
9			.46	-	.33	2.21	.190	5.08	.124	7.09
10					.41	1.78	.223	5.73	.141	8.16
11					.53	.70	.266	6.04	.162	9.04
12					.69	-	.313	6.47	.187	9.79
13							.378	6.34	.216	10.45
14							.461	5.87	.250	11.00
15							.57	4.92	.29	11.44
16							.73	2.80	.34	11.60
17							.91	1.24	.40	11.60
18							1.2	-	.49	10.52
19									.60	9.20
20									.72	8.32

4. Comments

1. The values $c_{n,p}$ are undoubtedly not the largest values of c for which δ^c is minimax. The approximations in the proof resulted in a less than optimal bound on c. It is generally desirable to choose $c = p - 2$ (see Berger (1980a) for indications of why this is so in the case of known $\ddagger$), so a good choice of c here is $c^* = \min\{p-2, c_{n,p}\}$. Luckily, as can be seen from Table 1, $p - 2$ is quite often less than $c_{n,p}$.

2. If $\mu = (\mu_1, \ldots, \mu_p)^t$ is thought to be the "most likely" value of θ, δ^c should shrink towards μ rather than zero. The appropriate estimator is then

$$\delta^c(X,W) = (I - c\alpha(Q^{\frac{1}{2}} WQ^{\frac{1}{2}})h([X-\mu]^t W^{-1}[X-\mu])Q^{-1}W^{-1})(X-\mu) + \mu .$$

This estimator is also minimax under the given conditions on c, α, and h, as a simple transformation shows.

More generally, if a linear restriction on θ is thought to hold, then δ^c can (and should) be modified to shrink towards the subspace determined by the linear restriction. (See Theorem 2 of Berger and Bock (1977) for the method of doing this.) The point is that δ^c, though uniformly better than δ^0, is significantly better only for θ near the region towards which the estimator shrinks. Hence it is crucial to shrink towards the region in which θ is thought to lie.

3. If $Q\ddagger$ has a very broad eigenvalue spectrum, then $\alpha(Q^{\frac{1}{2}} WQ^{\frac{1}{2}})h(X^t W^{-1} X) (\leq ch_{min}(Q^{\frac{1}{2}} WQ^{\frac{1}{2}})/[(X^t W^{-1} X)(n-p-1)])$ will tend to be small. From (3.7), it follows that Δ_c will be close to zero, and hence δ^c will improve little upon δ^0. If, therefore, certain coordinates of X are considered likely to result in comparatively small eigenvalues of $Q\ddagger$, it would pay to estimate those coordinates separately.

4. For known ‡, it is argued in Berger (1980a) and Berger (1980b) that in choosing an alternative to δ^0 a partly Bayesian attitude is necessary (for basically the reason given in Comment 2). Minimax results for appealing Bayesian estimators were given in Berger (1980a) and especially Berger (1980b). It would be very nice if similar results could be developed for unknown ‡, but that necessitates consideration of a much wider class of estimators than given by (1.1). Unfortunately, we were unable to establish minimax results for other classes of estimators. The techniques used here, particularly the employment of the identities of Haff (1980b), will probably be indispensable in any such generalization, however.

5. For an illustration of the amount of improvement obtainable by using δ^C, the reader is referred to Figure 1 in B^3CG(1977).

Acknowledgements: We are grateful to Regina Becker for assistance in the computer calculations.

References

[1] Alam, Kursheed (1975). Minimax and admissible minimax estimators of the mean of a multivariate normal distribution for unknown covariance matrix. J. Multivariate Anal. 5, 83-95.

[2] Berger, J. (1976). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. J. Multivariate Anal. 6, 256-264.

[3] Berger, J. (1980a). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. Ann. Statist. 8, 716-761.

[4] Berger, J. (1980b). Selecting a minimax estimator of a multivariate normal mean. Technical Report #80-24, Purdue University.

[5] Berger, J. and Bock, M. E. (1976a). Combining independent normal mean estimation problems with unknown variances. Ann. Statist. 4, 642-648.

[6] Berger, J. and Bock, M. E. (1976b). Eliminating singularities of Stein-type estimators of location vectors. J. Roy. Statist. Soc. Ser. B 38, 166-170.

[7] Berger, J. and Bock, M. E. (1977). Improved minimax estimators of normal mean vectors for certain types of covariance matrices. In Statistical Decision Theory and Related Topics II, S. S. Gupta and D. S. Moore (Eds.). Academic Press, New York.

[8] Berger, J., Bock, M. E., Brown, L. D., Casella, G., and Gleser, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. Ann. Statist. 5, 763-771.

[9] Bock, M. E. (1974). Certain minimax estimators of the mean of a multivariate normal distribution. Ph.D. thesis, Univ. of Illinois.

[10] Brown, P. J. and Zidek, J. V. (1979). Multivariate ridge regression with unknown covariance matrix. Technical Report No. 79-11, Mathematics Dept., Univ. of British Columbia.

[11] Dykstra, R. L. and Hewett, J. E. (1978). Positive dependence of the roots of a Wishart matrix. Ann. Statist. 6, 235-238.

[12] Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. Ann. Statist. 4, 11-21.

[13] Gleser, L. (1979). Minimax estimation of a normal mean vector when the covariance matrix is unknown. Ann. Statist. 7, 838-846.

[14] Haff, L. R. (1976). Minimax estimators of the multinormal mean: autoregressive priors. J. Multivariate Anal. 6, 265-280.

[15] Haff, L. R. (1977). Minimax estimators for a multinormal precision matrix. J. Multivariate Anal. 7, 374-385.

[16] Haff, L. R. (1979). An identity for the Wishart distribution with applications. J. Multivariate Anal. 9, 531-542.

[17] Haff, L. R. (1980a). Empirical Bayes estimation of the multivariate normal covariance matrix. Ann. Statist. 8, 586-597.

[18] Haff, L. R. (1980b). Identities for the inverse Wishart distribution and their relation to the Wishart case. Unpublished manuscript.

[19] James, W. and Stein, C. (1960). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1, 361-379, Univ. of California Press.

[20] Judge, G. and Bock, M. E. (1978). Statistical Implications of Pre-Test and Stein Rule Estimators in Econometrics. North Holland Publishing Company, Amsterdam.

[21] Kamae, T., Krengel, U., and O'Brien, G. L. (1977). Stochastic inequalities on partially ordered spaces. Ann. Prob. 5, 899-912.

[22] Lin, Pi-Erh and Tsai, Hui-Lang (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. Ann. Statist. 1, 142-145.

[23] Marshall, A. W. and Olkin, I. (1979). Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

[24] Shinozaki, N. (1977). Simultaneous estimation of the means of independent variables with unknown variances. Keio Math. Sem. Rep. 2, 75-79.

[25] Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob. 1, 197-206, Univ. of California Press.

[26] Stein, C. (1973). Estimation of the mean of a multivariate distribution. Proc. Prague Symp. Asymptotic Statist. 345-381.

[27] Veinott, A. F., Jr. (1965). Optimal policy in a dynamic, single product, non-stationary inventory model with several demand classes. Operations Res. 13, 761-778.

[28] Wilkinson, J. H. (1965). The Algebraic Eigenvalue Problem. Clarendon Press, Oxford.