New Estimators for the Mean Vector of a Normal Distribution with Unknown Covariance Matrix

by

Leon Jay Gleser
Purdue University

Technical Report #86-17

Department of Statistics
Purdue University

May 1986

NEW ESTIMATORS FOR THE MEAN VECTOR OF A NORMAL DISTRIBUTION WITH UNKNOWN COVARIANCE MATRIX

Leon Jay Gleser
Department of Statistics
Purdue University
West Lafayette, IN 47907

1. INTRODUCTION

A p-dimensional ($p \geq 3$) random column vector $X = (x_1, \ldots, x_p)'$ is observed, where X has a multivariate normal distribution with mean vector μ and positive definite covariance matrix Σ. Also observed is the $p \times p$ random matrix $W = ((w_{ij}))$, which is statistically independent of X and has a Wishart distribution with n degrees of freedom ($n > p+1$) and expected value $E(W) = n\Sigma$. It is desired to estimate μ by an estimator $\delta = \delta(X, W)$ under the quadratic loss.

$$L(\delta; \mu, \Sigma) = [\mathrm{tr}(Q\Sigma)]^{-1}(\delta - \mu)'Q(\delta - \mu), \tag{1.1}$$

where Q is a known positive-definite matrix and $\mathrm{tr}(A)$ stands for the trace of the matrix A.

The above problem is the canonical form of the situation (see, for example, Anderson 1984, Chapter 3) where i.i.d. observations $Y_1, Y_2, \ldots Y_N$ are taken from a p-dimensional normal population with mean vector μ and covariance matrix ψ, and the data are reduced by sufficiency to $X = N^{-1} \sum_{i=1}^{N} Y_i$, $W = N^{-1} \sum_{i=1}^{N} (Y_i - X)(Y_i - X)'$. Here, $\Sigma = N^{-1}\psi$, $n = N - 1$.

Regardless of whether or not Σ is known, and for any Q, the estimator X is minimax for μ. However, beginning with the landmark paper of Stein (1956), a large body of research has been devoted to establishing broad classes of estimators which dominate X in risk, often substantially. For the most part, such research has concentrated upon cases where Σ is known (in which case W is not needed), or known up to a positive scalar multiple σ^2.

The case of completely unknown Σ is more complex, and thus has been more resistant to solution. Berger *et al* (1977), Gleser (1979) and Berger and Haff (1983) have succeeded in developing estimators which dominate X in risk. However, only Berger and Haff (1983) have been able to provide a completely analytic proof, and this only for a subset of the class of estimators which they considered.

In Gleser (1986), a new method is given which yields (for the first time in the unknown-Σ context) an unbiased estimator of risk for certain classes of estimators. The present companion paper applies this method to construct two fairly broad classes of estimator which dominate X in risk. In both cases, completely analytic proofs of risk dominance against the estimator X are obtained.

In Section 2 the results in Gleser (1986) are briefly described. Section 3 gives risk-domination results (versus X) for a class of estimators closely related to the estimators considered by Berger and Haff (1983). Section 4 gives risk-domination results for a class of estimators motivated by estimators constructed by Berger (1976) in the known-Σ case.

2. UNBIASED ESTIMATOR OF RISK

To reduce notational complexity, in the remainder of this paper it is assumed that

$$Q = I_p, \tag{2.1}$$

where I_p is the p-dimensional identity matrix. Estimators for the general Q case can be obtained from estimators for the case (2.1) as follows:

$$\delta^*(X,W) = (T')^{-1}\, \delta(T'X, T'WT) \tag{2.2}$$

where T is any solution of $Q = TT'$. An estimator $\delta(X,W)$ dominates X in risk in the case (2.1) if and only if $\delta^*(X,W)$ defined by (2.2) dominates X in risk when the loss function (1.1) is defined by general Q.

Let

$$h(X,W) = (h_1(X,W), h_2(X,W), \ldots, h_p(X,W))'$$

be a p-dimensional vector-valued function of X and W. Let

$$\nabla h(X,W) = \left(\left(\frac{\partial h_i(X,W)}{\partial X_j}\right)\right), \tag{2.3}$$

and let $r(X,W) = (r_1(X,W), \ldots, r_p(X,W))'$ be defined by

$$r_i(X,W) = \frac{\partial(W\ h(X,W))_i}{\partial w_{ii}} + \frac{1}{2}\sum_{j\neq i} \frac{\partial(W\ h(X,W))_j}{\partial w_{ij}}, \quad i = 1,2,\ldots,p. \tag{2.4}$$

It is assumed that $h(X,W)$ is sufficiently regular as a function of X and W that the partial derivatives in (2.3) and (2.4) exist almost everywhere (in the product space of X and W), and further that $h(X,W)$ permits certain integration-by-parts identities for expected values over X and W to hold (see Gleser (1986) for further details).

THEOREM 1. *Let*

$$t(X,W) = h(X,W) + \frac{2}{n-p-1}\, r(X,W). \tag{2.5}$$

Then if $\delta(X,W) = X - t(X,W)$ has finite risk,

$$\mathrm{tr}(\Sigma)[R(\delta;\mu,\Sigma) - R(X;\mu,\Sigma)] = E[M(X,W)]$$

where

$$M(X,W) = t'(X,W)t(X,W) - \frac{2}{n-p-1}\,\mathrm{tr}(W\ \nabla h(X,W)), \tag{2.6}$$

and $R(\delta;\mu,\Sigma) = E[L(\delta(X,W);\mu,\Sigma)]$ *is the risk function for the estimator* $\delta = \delta(X,W)$.

Theorem 1 is proven in Gleser (1986). Note that the unbiased estimator $M(X,W)$ of risk difference obtained in Theorem 1 is for the estimator $X - t(X,W)$ rather than for $X - h(X,W)$. Use of Theorem 1 thus yields risk dominance proofs for modifications of an estimator $X - h(X,W)$, rather than the estimator itself. As will be seen in Section 3, this is not a severe handicap. Indeed, the modifications made to an estimator often are intuitively reasonable (and can even be thought of as improvements).

3. A CLASS OF MODIFIED ESTIMATORS

Berger and Haff (1983) consider estimators for μ of the form

$$\delta(X,W) = X - c\ \alpha(W)\ s(X'W^{-1}X)W^{-1}X, \tag{3.1}$$

where $c \geq 0$, the scalar function $s(\cdot)$ maps $[0,\infty)$ into $[0,\infty)$ and is continuous and piecewise differentiable (with derivative $s^{(1)}(z) = ds(z)/dz$), and the scalar function $\alpha(W)$ is everywhere nonnegative, continuous, and differentiable (with respect to the elements of W). The results given in this section concern modifications of this class of estimators.

Note: In comparing the results of the present paper with those of Berger and Haff (1983), note that here it is assumed that $Q = I_p$, so that the transformation (2.2) is needed to transform (3.1) into the general-Q estimators of Berger and Haff. Further, our $s(\cdot)$ is their $h(\cdot)$.

For notational convenience let

$$v = X'W^{-1}X.$$

Let

$$h(X,W) = c\ \alpha(W)s(X'W^{-1}X)W^{-1}X = c\ \alpha(W)s(v)W^{-1}X.$$

Then

$$\mathrm{tr}(W\nabla h(X,W)) = c\ \alpha(W)[2vs^{(1)}(v) + ps(v)]. \tag{3.3}$$

Recall that

$$\frac{\partial w^{ab}}{\partial w_{ij}} = \begin{cases} -w^{ai}w^{jb} - w^{aj}w^{ib}, & i \neq j, \\ -w^{ai}w^{ib}, & i = j, \end{cases} \tag{3.4}$$

where $W^{-1} = ((w^{ab}))$. Using (2.4), (3.3), the chain rule and straightforward algebra,

$$\begin{aligned} r(X,W) &= (r_1(X,W),\ldots,r_p(X,W))' \\ &= c\alpha(W)[s(v)U(X)X - vs^{(1)}(v)W^{-1}X], \end{aligned} \tag{3.5}$$

where

$$U(W) = ((u_{ij}(W))), \quad u_{ij}(W) = \begin{cases} \frac{\partial \log \alpha(W)}{\partial w_{ii}}, & i = j, \\ \frac{1}{2}\frac{\partial \log \alpha(W)}{\partial w_{ij}}, & i \neq j. \end{cases} \tag{3.6}$$

It thus follows from (2.5) and (3.5) that

$$t(X,W) = c\alpha(W)[(s(v) - \frac{2}{n-p-1}vs^{(1)}(v))W^{-1}X + \frac{2}{n-p-1}s(v)U(W)X]. \tag{3.7}$$

LEMMA 1. For any two p-dimensional column vectors $x, y (x \neq 0)$ and any two scalars c_1, c_2,

$$(c_1x + c_2y)'\,(c_1x + c_2y) \leq x'x\left\{|c_1| + |c_2|\left(\frac{y'y}{x'x}\right)^{\frac{1}{2}}\right\}^2 .$$

Proof. See Gleser (1986). □

LEMMA 2. Assume that

$$U'(W)U(W) \leq W^{-2}, \tag{3.8}$$

in the sense of the ordering of semi-definiteness for matrices. Assume also that

$$\alpha(W) \leq \frac{\lambda_{\min}(W)}{n-p-1}, \tag{3.9}$$

where $\lambda_{\min}(A)$ denotes the smallest characteristic root of a symmetric matrix A. Then

$$t'(X,W)t(X,W) \leq \left(\frac{c^2\alpha(W)}{n-p-1}\right) v\left\{|s(v) - \frac{2}{n-p-1}vs^{(1)}(v)| + \frac{2}{n-p-1}s(v)\right\}^2 .$$

Proof. It follows from (3.8) that $X'U'(W)U(W)X \leq X'W^{-2}X$. Also note that for any X

$$X'W^{-2}X \leq \frac{X'W^{-1}X}{\lambda_{\min}(W)}.$$

From these two facts, (3.7), (3.9), and Lemma 1 (with $x = W^{-1}X, y = U(W)X$), the assertion of Lemma 2 follows. □

If the risk of $\delta(X,W) = X - t(X,W)$ is finite, it follows from Theorem 1 that

$$M(X,W) = t'(X,W)t(X,W) - \frac{2}{n-p-1}\,\mathrm{tr}(W\ \nabla h(X,W))$$

is an unbiased estimator of the weighted risk difference

$$\mathrm{tr}(\Sigma)[R(\delta;\mu,\Sigma) - R(X;\mu,\Sigma)].$$

Consequently, for $\delta(X,W)$ to dominate X in risk it is sufficient that $M(X,W) \leq 0$. However, it follows from (3.3) and Lemma 2 that

$$M(X,W) \leq \left(\frac{c\alpha(W)}{n-p-1}\right) g(v), \tag{3.10}$$

where

$$\begin{aligned} g(v) = cv[|s(v) - \frac{2}{n-p-1} v s^{(1)}(v)| + \frac{2}{n-p-1} s(v)]^2 \\ - 2[2vs^{(1)}(v) + ps(v)]. \end{aligned} \tag{3.11}$$

THEOREM 2. *If $\delta(X,W) = X - t(X,W)$ has finite risk, the assumptions* (3.8), (3.9) *hold, and further*

$$g(v) \leq 0, \qquad \text{all } \ v \geq 0,$$

then $\delta(X,W)$ dominates X in risk.

Proof. Since it is given that $c \geq 0, \alpha(W) \geq 0$, and $n-p-1 \geq 0$, the assertion of Theorem 2 is a direct consequence of Theorem 1 and (3.10). □

COROLLARY 1. Suppose that $\alpha(W)$ satisfies (3.8), (3.9) and that $s(v)$ satisfies "Condition h" of Berger and Haff (1983); that is, $s(v)$ is continuous and piecewise differentiable and for all v,

$$0 \leq vs(v) \leq 1, \tag{3.12a}$$

$$s(v) + vs^{(1)}(v) \geq 0, \tag{3.12b}$$

$$s^{(1)}(v) \leq 0, \tag{3.12c}$$

Then $\delta(X,W)$ dominates X in risk if $\delta(X,W)$ has finite risk and

$$0 < c \leq 2(p-2)\left[\frac{n-p-1}{n-p+3}\right]^2.$$

Proof. It follows from (3.11) and (3.12c) that

$$g(v) = \frac{cv}{(n-p-1)^2}[(n-p+1)s(v) - 2vs^{(1)}(v)]^2 - 2[2vs^{(1)}(v) + ps(v)].$$

However, (3.12b) implies that

$$\begin{aligned} g(v) &\leq \frac{cv}{(n-p-1)^2}[(n-p+1)s(v) - 2vs^{(1)}(v)]^2 - 2(p-2)s(v) \\ &= s(v)\left\{\frac{cvs(v)}{(n-p-1)^2}\left[(n-p+1) - 2\left(\frac{vs^{(1)}(v)}{s(v)}\right)\right]^2 - 2(p-2)\right\}. \end{aligned}$$

Next, apply (3.12a) to obtain

$$g(v) \le s(v) \left\{ \frac{c}{(n-p-1)^2} \left[(n-p+1) - 2 \left[\frac{vs^{(1)}(v)}{s(v)} \right] \right]^2 - 2(p-2) \right\}.$$

However

$$\begin{aligned} \left[(n-p+1 - \frac{2vs^{(1)}(v)}{s(v)} \right]^2 &= (n-p+1)^2 - 4(n-p+1)\frac{vs^{(1)}(v)}{s(v)} + 4 \left[\frac{vs^{(1)}(v)}{s(v)} \right]^2 \\ &\le (n-p+1)^2 + 4(n-p+1) + 4 \\ &= (n-p+3)^2, \end{aligned}$$

since by (3.12b) and (3.12c),

$$0 \le \frac{-vs^{(1)}(v)}{s(v)} \le 1.$$

Thus

$$g(v) \le s(v) \left\{ \frac{c(n-p+3)^2}{(n-p-1)^2} - 2(p-2) \right\}.$$

The conclusion of Corollary 1 is now an immediate consequence of Theorem 2. □

Remark 1. The function $s(v) = v^{-1}$ satisfies the conditions of Corollary 1. This special case is also treated in Gleser [1986, Section 3].

Remark 2. Condition (3.9) on $\alpha(W)$ in this paper is identical to (i) of Berger and Haff's (1983) "Condition α". It is not clear how (3.8) in this paper relates to (ii) and (iv) of "Condition α" in Berger and Haff (1983); however, the examples

$$\alpha(W) = \frac{1}{n-p-1}\lambda_{\min}(W), \quad \alpha(W) = \frac{1}{(n-p-1)\mathrm{tr}(W^{-1})}$$

of possible choices for $\alpha(W)$ mentioned in Berger and Haff (1983) satisfy (3.8) and (3.9) and also "Condition α".

The main results of Berger and Haff (1983) and Theorem 2 and Corollary 1 of this paper are each concerned with finding bounds on a constant c which allow estimators of the form $X - cf(X,W)$ to dominate X in risk. However, the classes of estimators considered are not the same. In Berger and Haff (1983), $f(X,W)$ has the form

$$f(X,W) = \alpha(W)s(X'W^{-1}X)W^{-1}X, \tag{3.13}$$

while in the present paper this function is modified as follows:

$$\begin{aligned} f(X,W) = \alpha(W)[s(X^{-1}W^{-1}X) - \frac{2}{n-p-1}(X'W^{-1}X)s^{(1)}(X'W^{-1}X)]W^{-1}X \\ + \frac{2}{(n-p-1)}\alpha(W)s(X'W^{-1}X)U(W)X, \end{aligned} \tag{3.14}$$

where $U(W)$ is defined from $\alpha(W)$ by (3.6).

Nevertheless, these two classes are closely related. This is perhaps most easily seen by considering the special case $\alpha(W) = (n-p-1)^{-1}\lambda_{\min}(W), s(v) = v^{-1}$. Here, (3.13) becomes

$$\left[\frac{\lambda_{\min}(W)}{(n-p-1)X'W^{-1}X}\right] W^{-1}X, \tag{3.15}$$

while (3.14) is

$$\left(\frac{n-p+1}{n-p-1}\right)\frac{\lambda_{\min}(W)}{(n-p-1)X'W^{-1}X}\left\{W^{-1} + \left(\frac{2}{n-p+1}\right)\frac{1}{\lambda_{\min}(W)}g_p g_p'\right\}X, \tag{3.16}$$

where g_p is the normalized characteristic vector of W corresponding to $\lambda_{\min}(W)$. If $\lambda_1(W) \geq \ldots \geq \lambda_p(W) = \lambda_{\min}(W)$ are the ordered characteristic roots of W, and $g_1, \ldots, g_p$ the corresponding characteristic vectors, then

$$W^{-1} = \sum_{i=1}^{p} \frac{1}{\lambda_i(W)} g_i g_i',$$

while

$$W^{-1} + \frac{2}{n-p-1}\frac{1}{\lambda_{\min}(W)}\, g_p g_p' = \sum_{i=1}^{p-1} \frac{1}{\lambda_i(W)}\, g_i g_i' + \left(\frac{n-p+3}{n-p+1}\right)\frac{1}{\lambda_p(W)}\, g_p g_p'.$$

Thus, the estimators (3.15) and (3.16) are nearly the same, particularly if $n-p$ is large.

Therefore, a comparison of the maximum values of c permitting $\delta(X, W)$ to dominate X in risk given respectively by Corollary 1 of this section and Tables 1 and 2 of Berger and Haff (1983) can be informative. Berger and Haff actually give three different bounds for c:

(1) A general analytic bound $c^*_{n,p}$ given in Table 2 and Equation (1.3) in Berger and Haff (1983). This bound holds for all $\alpha(W), s(v)$ respectively satisfying "Condition α" and "Condition h" in their paper.

(2) A bound $c^{(1)}_{n,p}$ given by Corollary 3(a) and columns marked G in Table 1 of Berger and Haff (1983). This bound applies only to the case where $\alpha(W) = (n-p-1)^{-1}\lambda_{\min}(W)$ and is the *maximum* of bounds obtained by Monte Carlo simulation in Gleser (1979) and Berger and Haff (1983). (Standard errors for the simulated bounds are given by Berger and Haff (1983).) As Berger and Haff (1983) note, the bounds $c^{(1)}_{n,p}$ are typically considerably larger than the bounds $c^*_{n,p}$.

(3) A bound $c^{(2)}_{n,p}$ given by Corollary 3(b) and columns marked B in Table 1 of Berger and Haff (1983). These bounds are further specialized over the bounds $c^{(1)}_{n,p}$ in that they apply only to the case $s(v) = v^{-1}$. Again, these bounds are obtained

by simulation and are the maxima of bounds obtained by Berger *et al* (1977), Gleser (1979) and Berger and Haff (1983).

Table 1 compares the bounds $c^*_{n,p}, c^{(1)}_{n,p}, c^{(2)}_{n,p}$ to the bound $\tilde{c} = 2(p-2)[(n-p-1)/(n-p+3)]^2$ obtained from Corollary 1 of this paper. The bound $\tilde{c}$ applies to all of the contexts used above [except that, as noted earlier, conditions (3.8) and (3.9) of this paper may not be as general as "Condition α" in Berger and Haff's (1983) paper].

Table 1 reveals that the general analytic bounds $\tilde{c}$ of this paper are uniformly larger than the general analytic bounds $c^*_{n,p}$ in Berger and Haff (1983), suggesting that the analytic methods in Gleser (1986) may be more sensitive than those of Berger and Haff (1983). [Note: A blank entry in Table 1 indicates that the corresponding bound is not positive, and thus useless.] Further, $\tilde{c}$ is larger than the more specialized bound $c^{(1)}_{n,p}$ when $n-p$ is either small or very large. However, the bound $c^{(2)}_{n,p}$ is generally larger than $\tilde{c}$. These observations indicate that the Monte Carlo simulation approach has an advantage over the purely analytic methods of this paper and of Berger and Haff (1983), particularly for finding maximal allowable "shrinkage" for narrowly specified classes of estimators (e.g., when $\alpha(W)$ and $s(v)$ are specified). This, of course, is not surprising since analytic arguments of necessity (because of their generality) must use fairly crude inequalities. (See, for example, the proof of Corollary 1 in this section.) Perhaps, the real surprise is that the analytic arguments in this paper do so well!

4. A NEW CLASS OF ESTIMATORS

For the case where Σ is known, Berger (1976) considered a class of estimators of μ which take the form (when $Q = I_p$):

$$X - \frac{u(X'\Sigma^{-2}X)}{X'\Sigma^{-2}X}\Sigma^{-1}X, \qquad u(\cdot) \text{ maps } [0,\infty) \text{ to } [0,\infty).$$

He shows that certain members of this class both dominate X in risk, and are themselves admissible. Consider the simplest member of the above class (where $u(\cdot) = c$, a constant), and substitute an estimator for Σ of the form $\alpha(W)W$, where $\alpha(W)$ is a nonnegative scalar function of W, for Σ. This yields an estimator of the form:

$$X - \left[\frac{c\alpha(W)}{X'W^{-2}X}\right] W^{-1}X. \tag{4.1}$$

Let

$$h(X,W) = \left[\frac{c\alpha(W)}{X'W^{-2}X}\right] W^{-1}X,$$

and apply the theory of Section 2.

Thus,

$$\mathrm{tr}(W\ \nabla h(X,W)) = \frac{c\alpha(W)(p-2)}{X'W^{-2}X}. \tag{4.2}$$

Table 1. Maximum Values of c

p	$n=16$: $\tilde{c}$	$c^*_{n,p}$	$c^{(1)}_{n,p}$	$c^{(2)}_{n,p}$	$n=18$: $\tilde{c}$	$c^*_{n,p}$	$c^{(1)}_{n,p}$	$c^{(2)}_{n,p}$	$n=20$: $\tilde{c}$	$c^*_{n,p}$	$c^{(1)}_{n,p}$	$c^{(2)}_{n,p}$
3	1.13	–	0.58	1.06	1.21	–	0.75	1.20	1.28	–	0.87	1.34
4	2.15	–	1.79	2.48	2.34	–	2.02	2.65	2.49	0.25	2.25	2.85
5	3.06	–	2.78	3.80	3.38	0.37	3.17	4.05	3.63	0.98	3.63	4.35
6	3.83	–	3.47	4.81	4.30	0.82	4.17	5.33	4.68	1.74	4.87	5.73
7	4.44	–	3.93	5.78	5.10	1.04	5.05	6.42	5.63	2.18	5.96	6.99
8	4.86	–	4.19	6.57	5.75	0.60	5.12	7.64	6.45	2.43	6.81	8.19
9	5.04	–	3.86	7.02	6.22	–	5.56	8.40	7.14	2.21	7.38	9.22
10	4.94	–	3.66	6.79	6.48	–	5.17	8.90	7.67	1.78	8.80	10.25
11	4.50	–	1.28	5.78	6.48	–	5.18	9.15	8.00	0.70	7.10	10.84
12	3.67	–	–	2.73	6.17	–	4.21	8.42	8.10	–	6.52	11.10
13	2.44	–	–	–	5.50	–	0.94	7.11	7.92	–	6.25	11.09
14	0.96	–	–	–	4.41	–	–	2.43	7.41	–	4.58	9.79
15	–	–	–	–	2.89	–	–	–	6.50	–	–	7.93
16	–	–	–	–	1.12	–	–	–	5.14	–	–	2.26
17	–	–	–	–	–	–	–	–	3.33	–	–	–
18	–	–	–	–	–	–	–	–	1.28	–	–	–
19	–	–	–	–	–	–	–	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–	–	–	–

p	$n=25$: $\tilde{c}$	$c^*_{n,p}$	$c^{(1)}_{n,p}$	$c^{(2)}_{n,p}$	$n=30$: $\tilde{c}$	$c^*_{n,p}$	$c^{(1)}_{n,p}$	$c^{(2)}_{n,p}$
3	1.41	–	1.20	1.51	1.50	–	1.27	1.59
4	2.78	0.61	2.78	3.09	2.97	1.20	2.98	3.33
5	4.09	1.60	4.28	4.79	4.41	2.41	4.71	5.09
6	5.36	2.56	5.85	6.43	5.81	3.60	6.38	6.80
7	6.55	3.51	7.29	7.93	7.16	4.69	8.00	8.47
8	7.68	4.26	8.58	9.26	8.47	5.88	9.60	10.15
9	8.73	5.08	9.76	10.60	9.72	7.09	11.28	11.80
10	9.68	5.73	10.92	11.98	10.92	8.16	12.62	13.37
11	10.53	6.04	11.85	13.14	12.05	9.04	14.10	14.74
12	11.25	6.47	12.66	14.20	13.11	9.79	15.35	16.06
13	11.83	6.34	12.66	15.48	14.08	10.45	16.40	17.36
14	12.24	5.87	12.41	15.74	14.96	11.00	17.53	18.72
15	12.46	4.92	11.77	16.61	15.73	11.44	18.57	19.90
16	12.44	2.80	10.91	16.67	16.37	11.60	18.78	20.62
17	12.15	1.24	10.45	16.67	16.88	11.60	19.00	21.56
18	11.52	–	9.30	16.34	17.21	10.52	18.11	22.38
19	10.49	–	–	–	17.35	9.20	14.84	22.83
20	9.00	–	–	–	17.25	8.32	14.52	23.47

Also, from (2.4),

$$r(X,W) = h(X,W) + \frac{c\alpha(W)X'W^{-1}X}{(X'W^{-2}X)^2}W^{-2}X + \frac{c\alpha(W)}{X'W^{-2}X}U(W)X,$$

where $U(W)$ is defined by (3.6). It then follows from (2.5) that

$$t(X,W) = c\left(\frac{n-p+1}{n-p-1}\right)\frac{\alpha(W)}{X'W^{-2}X}\Bigg\{W^{-1}X + \left(\frac{2}{n-p+1}\right)\left[\frac{X'W^{-1}X}{X'W^{-2}X}W^{-2}X + U(W)X\right]\Bigg\}. \tag{4.3}$$

By Theorem 1 and (4.2), the unbiased estimator of the weighted risk difference

$$(\mathrm{tr}(\Sigma))[R(\delta;\mu,\Sigma) - R(X;\mu,\Sigma)]$$

between

$$\delta(X,W) = X - t(X,W) \tag{4.4}$$

and X is

$$M(X,W) = t'(X,W)t(X,W) - \left(\frac{2c}{n-p-1}\right)\left(\frac{\alpha(W)}{X'W^{-2}X}\right)(p-2). \tag{4.5}$$

THEOREM 3. *Let $\alpha(W)$ be a nonnegative, continuous and differentiable function of W which satisfies*

$$U'(W)U(W) \le 4W^{-2}, \tag{4.6}$$

$$\alpha(W) \le [\lambda_{\min}(W)/\lambda_{\max}(W)]^2, \tag{4.7}$$

where $\lambda_{\max}(W), \lambda_{\min}(W)$ respectively are the largest and smallest characteristic roots of W, and $U(W)$ is defined from $\alpha(W)$ by (3.6). Then, $\delta(X,W)$ defined by (4.3) and (4.4) dominates X in risk if

$$0 < c \le 2(p-2)\frac{(n-p-1)}{(n-p+7)^2}. \tag{4.8}$$

Proof. Note that

$$t'(X,W)t(X,W) = c^2\left(\frac{n-p+1}{n-p-1}\right)^2\frac{\alpha^2(W)}{X'W^{-2}X}H(X,W), \tag{4.9}$$

where

$$H(X,W) = 1 + \frac{4}{(n-p+1)^2}\Bigg\{\frac{(X'W^{-1}X)^2(X'W^{-4}X)}{(X'W^{-2}X)^3} + \frac{(X'W^{-3}X)(X'W^{-1}X)}{(X'W^{-2}X)^2}(n-p+1) + \frac{X'U'(W)U(W)X}{X'W^{-2}X} + \frac{X'W^{-1}U(W)X}{X'W^{-2}X}(n-p+1) \; 2\frac{(X'W^{-1}X)(X'W^{-2}U(W)X)}{(X'W^{-2}X)^2}\Bigg\}.$$

By the Cauchy-Schwarz inequality,

$$X'W^{-a}U(W)X \leq (X'W^{-2a}X)^{1/2}(X'U'(W)U(W)X)^{1/2}$$

for $a = 1, 2, \ldots$. Further, since $\alpha(W)$ satisfies (4.5),

$$X'U'(W)U(W)X \leq 4X'W^{-2}X.$$

Hence,

$$H(X,W) \leq \left(\frac{n-p+5}{n-p+1}\right)^2 + \frac{4}{(n-p+1)^2}\{A_{1,2,4}(X,W) + A_{1,2,3}(X,W)(n-p+1) + 4A^{\frac{1}{2}}_{1,2,4}(X,W)\},$$

where

$$A_{i,j,k}(X,W) = \frac{(X'W^{-i}X)^{k-j}(X'W^{-k}X)^{j-i}}{(X'W^{-j}X)^{k-i}},$$

and $i \leq j \leq k$ are nonnegative integers.

Let

$$\gamma = \gamma(W) = \frac{\lambda_{\max}(W)}{\lambda_{\min}(W)}.$$

Then, Marshall and Olkin (1964) have shown that

$$A_{i,j,k}(X,W) \leq \left(\frac{\gamma^{k-i}-1}{k-i}\right)^{k-i}\left(\frac{\gamma^{j-i}-1}{j-i}\right)^{-(j-i)}\left(\frac{\gamma^{k-i}-1}{k-i}\right)^{-(k-i)}.$$

Hence,

$$A_{1,2,3}(X,W) \leq \frac{(\gamma+1)^2}{4\gamma}, \quad A_{1,2,4}(X,W) \leq \frac{4}{27}\frac{(\gamma^2+\gamma+1)^3}{[\gamma(\gamma+1)]^2},$$

and since $\gamma \geq 1$,

$$\begin{aligned} H(X,W) &\leq \left(\frac{n-p+5}{n-p-1}\right)^2 + \frac{4}{(n-p+1)^2}\left\{(n-p+1)\frac{(\gamma+1)^2}{4\gamma} + \frac{4}{27}\frac{(\gamma^2+\gamma+1)^3}{[\gamma(\gamma+1)]^2} + \frac{8\left(\frac{1}{3}(\gamma^2+\gamma+1)\right)^{\frac{3}{2}}}{\gamma(\gamma+1)}\right\} \\ &\leq \left(\frac{n-p+5}{n-p-1}\right)^2 + \frac{4\gamma^2}{(n-p+1)^2}\{(n-p+1)+1+4\}. \end{aligned}$$

It now follows from (4.5), (4.7), (4.9), the definition of $\gamma = \gamma(W)$, and the fact that $\gamma(W) \geq 1$, that

$$\begin{aligned} M(X,W) &\leq \frac{c\alpha(W)}{X'W^{-2}X}\left\{c\alpha(W)\left[\left(\frac{n-p+5}{n-p-1}\right)^2 + \frac{4\gamma^2(n-p+6)}{(n-p-1)^2}\right] - \frac{2(p-2)}{n-p-1}\right\} \\ &\leq \frac{c\alpha(W)}{X'W^{-2}X}\left\{c\left(\frac{n-p+7}{n-p-1}\right)^2 - \frac{2(p-2)}{n-p-1}\right\}. \end{aligned}$$

The assertion of Theorem 3 now immediately follows. □

To show that there exist functions $\alpha(W)$ for which both (4.6) and (4.7) hold, note that for $\alpha(W) = \gamma^{-2}(W)$, which trivially satisfies (4.7),

$$U(W) = 2\left(\frac{1}{\lambda_{\min}(W)}\, g_p g_p' - \frac{1}{\lambda_{\max}(W)}\, g_1 g_1'\right),$$

where g_1, g_p are the normed characteristic vectors of W corresponding to $\lambda_{\max}(W)$, $\lambda_{\min}(W)$, respectively. Thus,

$$\begin{aligned} U'(W)U(W) &= 4\left[\frac{1}{\lambda^2_{\min}(W)}\, g_p g_p' + \frac{1}{\lambda_{\max}}\, g_1 g_1'\right] \\ &\leq 4W^{-2}, \end{aligned}$$

so that (4.6) is satisfied.

Note that $\gamma(W)$ is the condition number of W. Since the condition number of a matrix is related to the numerical stability of W^{-1} (or W^{-2}), it is not unreasonable to find $\gamma(W)$ appearing in the assumptions (and proof) of Theorem 3.

The analysis and form of the estimators in this section illustrate a major drawback to use of Theorem 1. In order to obtain an unbiased estimator of risk, it was necessary to switch from the estimator (4.1) to the considerably more complicated estimator $X - t(X, W)$. It would thus be highly desirable to obtain a refinement of Theorem 1 that does not require this switch. Some comments on this problem appear in Gleser (1986).

5. CONCLUSION

Two new classes of estimators which dominate X in risk have been given in this paper. Although the estimators, and method of analysis, are still somewhat crude, the results are still a great improvement over previous work, and suggest that Theorem 1 can play a useful role in developing applicable minimax estimators for μ.

ACKNOWLEDGEMENT

This research was supported by National Science Foundation Grant MCS-8501966.

REFERENCES

[1] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (2nd edition). New York: John Wiley & Sons.

[2] Berger, James O. (1986). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* *4*, 223–226.

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

[4] Berger, J. and Haff, L. R. (1983). A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Statistics and Decisions 1*, 105–129.

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

[6] Gleser, L. J. (1986). Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* *14*, to appear.

[7] Marshall, A. W. and Olkin, I. (1964). Reversal of the Lyapunov, Hölder and Minkowski inequalities and other extensions of the Kantorovich inequality. *J. Math. Anal. and Appl.* *8*, 503–514.

[8] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probability 1*, 197–206.