The Annals of Statistics
1979, Vol. 7, No. 6, 1264–1276

ESTIMATION OF THE INVERSE COVARIANCE MATRIX: RANDOM MIXTURES OF THE INVERSE WISHART MATRIX AND THE IDENTITY

BY L. R. HAFF

University of California, San Diego

Let $S_{p \times p}$ have a nonsingular Wishart distribution with unknown matrix Σ and k degrees of freedom. For two different loss functions, estimators of Σ^{-1} are given which dominate the obvious estimators aS^{-1}, $0 < a \leq k - p - 1$. Our class of estimators $\mathcal{C}$ includes random mixtures of S^{-1} and I. A subclass $\mathcal{C}_0 \subset \mathcal{C}$ was given by Haff. Here, we show that any member of $\mathcal{C}_0$ is dominated in $\mathcal{C}$. Some troublesome aspects of the estimation problem are discussed, and the theory is supplemented by simulation results.

1. Introduction and summary. Let $\mathbf{S}_{p \times p}$ have a nonsingular Wishart distribution with unknown matrix Σ and k degrees of freedom; i.e.,

$$\mathbf{S} \sim W(\Sigma, k), \qquad k - p - 1 > 0. \tag{1.1}$$

This paper is concerned with the problem of determining good estimates of Σ^{-1}. The results given constitute a generalization of those in [4] and are obtained using the same method. Some peculiar and troublesome aspects of the problem are also discussed.

Our estimators of Σ^{-1} have the form

$$\hat{\Sigma}^{-1} = [a + f(\mathbf{S})]\mathbf{S}^{-1} + g(\mathbf{S})I, \tag{1.2}$$

$0 < a \leq k - p - 1$, where $f(\mathbf{S})$ and $g(\mathbf{S})$ are real. In an earlier paper (Haff [4]) we gave estimators with $f(\mathbf{S}) \equiv 0$. It is seen, below, that such estimators are dominated by others with $f(\mathbf{S}) \not\equiv 0$.

Two different loss functions L_1 and L_2 are considered where

$$L_1(\hat{\Sigma}^{-1}, \Sigma^{-1}) = \operatorname{tr}(\hat{\Sigma}^{-1} - \Sigma^{-1})^2\mathbf{S}, \tag{1.3}$$

and

$$L_2(\hat{\Sigma}^{-1}, \Sigma^{-1}) = \operatorname{tr}(\hat{\Sigma}^{-1} - \Sigma^{-1})^2 Q,$$

where Q is an arbitrary p.d. matrix. Efron and Morris [2] developed L_1 from an empirical Bayes argument. Note that L_2, a more typical loss function, is the squared Euclidean norm of $\hat{\Sigma}^{-1} - \Sigma^{-1}$ when $Q = I$.

We shall evaluate an estimator in terms of its risk function

$$R_i(\hat{\Sigma}^{-1}, \Sigma^{-1}) \equiv EL_i(\hat{\Sigma}^{-1}, \Sigma^{-1}), \qquad i = 1 \text{ or } 2, \tag{1.4}$$

Received November 1977; revised August 1978.

AMS 1970 *subject classifications.* Primary 62F10; secondary 62C99.

Key words and phrases. Inverse covariance matrix, Stokes' theorem, integration by parts, Stein-like estimators, quadratic loss.

the latter being an average with respect to (1.1). As usual, if $\hat{\Sigma}_*^{-1}$ and $\hat{\Sigma}^{-1}$ are competing estimators of Σ^{-1}, then "$\hat{\Sigma}_*^{-1}$ dominates $\hat{\Sigma}^{-1}$ (mod L_i)" if $R_i(\hat{\Sigma}_*^{-1}, \Sigma^{-1}) < R_i(\hat{\Sigma}^{-1}, \Sigma^{-1})$ for all Σ.

The obvious estimators of Σ^{-1} are

$$\hat{\Sigma}_a^{-1} = a\mathbf{S}^{-1}, \qquad 0 < a \leqslant k - p - 1. \tag{1.5}$$

In particular, $E\hat{\Sigma}_{k-p-1}^{-1} = \Sigma^{-1}$. Efron and Morris [2] proposed estimators of the form

$$\hat{\Sigma}_{ab}^{-1} = a\mathbf{S}^{-1} + (b/\mathrm{tr}\,\mathbf{S})I \tag{1.6}$$

($f(\mathbf{S}) \equiv 0$, $g(\mathbf{S}) = b/\mathrm{tr}\,\mathbf{S}$, $b \geqslant 0$), and they proved $\hat{\Sigma}_{ab}^{-1}$ dominates $\hat{\Sigma}_a^{-1}$ (mod L_1) if $a = k - p - 1$ and $b = p^2 + p - 2$. (See, also, Stein, Efron and Morris [10].) Haff [4] proposed a larger class,

$$\hat{\Sigma}_*^{-1} = a\mathbf{S}^{-1} + [bt(U)/\mathrm{tr}\,\mathbf{S}]I \tag{1.7}$$

($f(\mathbf{S}) \equiv 0$, $g(\mathbf{S}) = bt(U)/\mathrm{tr}\,\mathbf{S}$, $U = p|\mathbf{S}|^{1/p}/\mathrm{tr}\,\mathbf{S}$), and obtained conditions under which

$$R_i(\hat{\Sigma}_*^{-1}, \Sigma^{-1}) < R_i(\hat{\Sigma}_{ab}^{-1}, \Sigma^{-1}) \leqslant R_i(\hat{\Sigma}_a^{-1}, \Sigma^{-1}) \qquad (\forall \Sigma), \qquad i = 1, 2. \tag{1.8}$$

In the present paper, we proceed as follows:

(a) We exploit an identity which was proved by Haff [4], viz,

$$\int_{\mathbb{S}} \mathrm{tr}[(\partial\phi/\partial S) \cdot F]\,dS = \int_{\mathbb{S}} \phi D^* F\,dS \tag{1.9}$$

where $\mathbb{S}$ denotes the cone of p.s.d. matrices, $F(S)$ a symmetric matrix, $\phi(S)$ a scalar function, and $D^*F \equiv \Sigma_{i \leqslant j}\partial F_{ij}/\partial s_{ij}$. Equation (1.9) was used to obtain identities for R_i, $i = 1, 2$ and to prove the dominance results (e.g., (1.8)). The same technique is used in the present paper. Here, however, F is nonsymmetric, so D^* is defined by

$$D^*F \equiv \Sigma\Sigma\partial F_{ij}/\partial s_{ij}. \tag{1.10}$$

Given such an F, as needed in Lemma 2.1, and given equation (1.10), the reader can verify that (1.9) remains valid by straightforward modification of [4], pages 377–379. We shall omit most of the details.

(b) We note some pathological results associated with L_1 and L_2. First, let us require $k - p - 3 > 0$; otherwise, $R_2(\hat{\Sigma}_a^{-1}, \Sigma^{-1})$ does not exist. For $k - p - 5 \leqslant a < k - p - 1$, it is seen that $\hat{\Sigma}_a^{-1}$ dominates $\hat{\Sigma}_{k-p-1}^{-1}$ (mod L_2); however, $\hat{\Sigma}_{k-p-1}^{-1}$ dominates $\hat{\Sigma}_a^{-1}$ (mod L_1), $a \neq k - p - 1$. The dominance reversal is surprising because $R_1(\hat{\Sigma}_a^{-1}, \Sigma^{-1})$ and $R_2(\hat{\Sigma}_a^{-1}, \Sigma^{-1})$ are qualitatively close if k is large and $Q = k\Sigma$. See Haff [4] for a similar anomaly associated with (1.6). (See, also, Efron and Morris [3] for a separate proof of the L_1 result.) Finally, we note the difficulty associated with $k - p = 4$ and $k - p = 5$. According to the L_2 result, $-\mathbf{S}^{-1}$ dominates $3S^{-1}(k - p - 1 = 3)$, and $0_{p \times p}$ dominates $4S^{-1}(k - p - 1 = 4)$!

(c) General conditions are given under which $\hat{\Sigma}^{-1} = [a + f(\mathbf{S})]\mathbf{S}^{-1}$ dominates

$\hat{\Sigma}_a^{-1}$(mod L_2). A special case is

$$\hat{\Sigma}^{-1} = \left[1 - c_1/\left(c_2 + |a\mathbf{S}^{-1}|\right)\right]\left[a\mathbf{S}^{-1}\right] \tag{1.11}$$

where c_i, $i = 1, 2$ are constants. Our generalization of (1.11) is similar in spirit to that of Efron and Morris [2] who generalized Stein's estimator for the multinormal mean.

(d) We show that any estimator with $f(\mathbf{S}) \equiv 0$ (e.g., (1.7)) can be dominated (mod L_2) by one from the class (1.2)—see Section 6.

(e) We present estimators of the form $\hat{\Sigma}^{-1} = a[1 - t(U)]\mathbf{S}^{-1} + [bt(U)/\mathrm{tr}\ \mathbf{S}]I$, $t(U) \nearrow$, $0 \leqslant t(U) \leqslant 1$, which dominate $\hat{\Sigma}_a^{-1}$(mod L_2). Here, U is given by $U_1 = p|\mathbf{S}|^{1/p}/\mathrm{tr}\ \mathbf{S}$ or $U_2 = p^2/[(\mathrm{tr}\ \mathbf{S}^{-1})(\mathrm{tr}\ \mathbf{S})]$. These ratios measure disparity among the sample eigenvalues; U_1 is the geometric mean over the arithmetic mean, U_2 the harmonic mean over the arithmetic mean

$$0 \leqslant U_2 \leqslant U_1 \leqslant 1.$$

If we assume that $\Sigma = \sigma^2 I$, then $(b/\mathrm{tr}\ \mathbf{S})I$, $b > 0$, are natural estimators of Σ^{-1}($b = pk - 2$ gives the unbiased estimator). An outcome $U_i \approx 1$ suggests the possibility $\Sigma^{-1} = (1/\sigma)^2 I$. Note that $\hat{\Sigma}^{-1}$ shifts weight from $a\mathbf{S}^{-1}$ to $(b/\mathrm{tr}\ \mathbf{S})I$ as $U_i \nearrow 1$.

The theory is supplemented by simulation results on each of the estimators (Section 7).

2. The risk identities. Let $\hat{\Sigma}^{-1}$ and $\hat{\Sigma}_a^{-1}$ be given by (1.2) and (1.5) respectively, $a \leqslant k - p - 1$. From (1.3) we have

$$R_1(\hat{\Sigma}^{-1}, \Sigma^{-1}) = R_1(\hat{\Sigma}_a^{-1}, \Sigma^{-1}) + \alpha_1(\Sigma)$$

and

$$R_2(\hat{\Sigma}^{-1}, \Sigma^{-1}) = R_2(\hat{\Sigma}_a^{-1}, \Sigma^{-1}) + \alpha_2(\Sigma)$$

where

$$\alpha_1(\Sigma) = E\left[(2af + f^2)\mathrm{tr}\ \mathbf{S}^{-1} + 2pg(a + f) + g^2\mathrm{tr}\ \mathbf{S} - 2f\ \mathrm{tr}\ \Sigma^{-1} - 2g\ \mathrm{tr}\ \Sigma^{-1}\mathbf{S}\right], \tag{2.1}$$

and

$$\alpha_2(\Sigma) = E\left[2af\ \mathrm{tr}(\mathbf{S}^{-2}Q) + 2ag\ \mathrm{tr}(\mathbf{S}^{-1}Q) + \mathrm{tr}(f\mathbf{S}^{-1} + gI)^2 Q - 2f\ \mathrm{tr}(\mathbf{S}^{-1}Q\Sigma^{-1}) - 2g\ \mathrm{tr}(\Sigma^{-1}Q)\right]. \tag{2.2}$$

To obtain (2.2), note that $\mathrm{tr}(\Sigma^{-1}S^{-1}Q) = \mathrm{tr}(\Sigma^{-1}S^{-1}Q)^t = \mathrm{tr}(QS^{-1}\Sigma^{-1}) = \mathrm{tr}(S^{-1}\Sigma^{-1}Q)$.

In this section, we give unbiased estimators of $\alpha_i(\Sigma)$, $i = 1, 2$, i.e., functions $\hat{\alpha}_i(\mathbf{S})$ for which $E\hat{\alpha}_i(\mathbf{S}) = \alpha_i(\Sigma)$, $i = 1, 2$. If $\hat{\alpha}_i(\mathbf{S}) \leqslant 0$ ($\forall S$), then "$\hat{\Sigma}^{-1}$ dominates $\hat{\Sigma}_a^{-1}$ (mod L_i)," $i = 1, 2$. Sufficient conditions for $\hat{\alpha}_i(\mathbf{S}) \leqslant 0$ are given in Sections 3, 4 and 6.

As in [4], we need the following operation: for a matrix M and real $t \neq 0$, we define $M_{(t)} \equiv tM + (1 - t)\mathrm{diag}(M)$ where $\mathrm{diag}(M)$ is the diagonal matrix with diagonal elements equal to those of M. When matrix inversion precedes (t), i.e., $(M^{-1})_{(t)}$, we simply write $M_{(t)}^{-1}$ because these operations are not used in the reverse order. It will be useful to note that $\mathrm{tr}[M_{(t)}N_{(1/t)}] = \mathrm{tr}(MN)$ and $\mathrm{tr}[M_{(t)}N] = \mathrm{tr}[MN_{(t)}]$ for matrices M and N.

The following lemma shows how to handle those terms in (2.1) and (2.2) in which the trace (under the expectation) explicitly depends on Σ.

LEMMA 2.1. *Let $f(\mathbf{S})$ and $g(\mathbf{S})$ satisfy the conditions of Stokes' theorem; also, let $f(\mathbf{S})$ and $g(\mathbf{S})$ $\|\mathbf{S}\|$ be bounded on $\mathbb{S}$.*

(a) *If $pk > 4$, then*

$$E\left[f(\mathbf{S})\mathrm{tr}\,\Sigma^{-1}\right] = \mathrm{tr}\,E\left[(k-p-1)f(\mathbf{S})\mathbf{S}^{-1} + 2(\partial f(\mathbf{S})/\partial S)\right]$$

and

$$E\left[g(\mathbf{S})\mathrm{tr}\,\Sigma^{-1}Q\right] = \mathrm{tr}\,E\left[(k-p-1)g(\mathbf{S})\mathbf{S}^{-1}Q + 2(\partial g(\mathbf{S})/\partial S)\cdot Q_{(\frac{1}{2})}\right].$$

(b) *If $pk > 2$, then*

$$E\left[g(\mathbf{S})\mathrm{tr}\,\Sigma^{-1}\mathbf{S}\right] = E\left[pkg(\mathbf{S}) + 2\,\mathrm{tr}\left(\partial g(\mathbf{S})/\partial S\cdot \mathbf{S}_{(\frac{1}{2})}\right)\right].$$

(c) *If $pk > 4$, then*

$$E\left[f(\mathbf{S})\mathrm{tr}\,\mathbf{S}^{-1}Q\Sigma^{-1}\right] = E\left\{(k-p-2)f(\mathbf{S})\mathrm{tr}(\mathbf{S}^{-2}Q) - f(\mathbf{S})(\mathrm{tr}\,\mathbf{S}^{-1})(\mathrm{tr}\,\mathbf{S}^{-1}Q) + 2\,\mathrm{tr}\left[(\partial f(\mathbf{S})/\partial S)\cdot(\mathbf{S}^{-1}Q)_{(\frac{1}{2})}\right]\right\}.$$

PROOF. These expectations are verified by straightforward modification of [4], pages 377–379. We omit the details. □

Finally, the unbiased estimators of $\alpha_i(\Sigma)$, $i = 1, 2$, are given by

THEOREM 2.1. *Let $f(\mathbf{S})$ and $g(\mathbf{S})$ satisfy the conditions of Lemma 2.1.*

(a) *For $a = a_0 = k - p - 1$,*

$$\hat{\alpha}_1(\mathbf{S}) = f^2\mathrm{tr}\,\mathbf{S}^{-1} - 4\,\mathrm{tr}\left(\frac{\partial f}{\partial S}\right) - 4\,\mathrm{tr}\left(\frac{\partial g}{\partial S}\right)\cdot \mathbf{S}_{(\frac{1}{2})} + 2pg(f-p-1) + g^2\mathrm{tr}\,\mathbf{S}.$$

(b) *For $a^* = a - a_0 \leqslant 0$,*

$$\begin{aligned}\hat{\alpha}_2(\mathbf{S}) = {} & f(f + 2a^* + 2)\mathrm{tr}(\mathbf{S}^{-2}Q) + 2g(a^* + f)\mathrm{tr}(\mathbf{S}^{-1}Q)\\ & -4\,\mathrm{tr}\left(\frac{\partial g}{\partial S}\right)\cdot Q_{(\frac{1}{2})} - 4\,\mathrm{tr}\left(\frac{\partial f}{\partial S}\right)\cdot T_{(\frac{1}{2})}\\ & +2f(\mathrm{tr}\,\mathbf{S}^{-1})(\mathrm{tr}\,\mathbf{S}^{-1}Q)\\ & +g^2\mathrm{tr}\,Q.\end{aligned}$$

PROOF. Immediate from Lemma 2.1. □

In the sequel, f and g depend only upon the eigenvalues of S; hence, the matrix $\partial f/\partial S$ (or $\partial g/\partial S$) is generally determined by standard perturbation theory. Let R

diagonalize S; i.e., $S = R\Lambda R'$, $\Lambda = \operatorname{diag}(l_1, \cdots, l_p)$ (the l_i distinct a.e.). Further, let $R = (R_1, R_2, \cdots, R_p)$, the eigenvectors given by $R_t = (r_{1t}, r_{2t}, \cdots, r_{pt})'$, $t = 1, \cdots, p$. For future reference, we record a standard result

$$\frac{\partial f(l_1, \cdots, l_p)}{\partial s_{ij}} = \Sigma_{k=1}^{p}\left(\frac{\partial f}{\partial l_k}\right)\left(\frac{\partial l_k}{\partial s_{ij}}\right) = \begin{cases} \Sigma_{k=1}^{p} r_{ik}^2\left(\dfrac{\partial f}{\partial l_k}\right) & i = j \\ 2\Sigma_{k=1}^{p} r_{ik}r_{jk}\left(\dfrac{\partial f}{\partial l_k}\right) & i \neq j. \end{cases} \tag{2.3}$$

3. The estimators $a\mathbf{S}^{-1}$. Let $a_0 = k - p - 1$, $f(\mathbf{S}) \equiv d$ (a nonnegative constant), and $g(\mathbf{S}) \equiv 0$. Here we compare

$$\hat{\Sigma}_a^{-1} = a\mathbf{S}^{-1}, \qquad a = a_0 - d, \tag{3.1}$$

with the unbiased estimator

$$\hat{\Sigma}_{a_0}^{-1} = a_0\mathbf{S}^{-1}. \tag{3.2}$$

From Press ([7], page 112) it can be seen that $R_2(\hat{\Sigma}_a^{-1}, \Sigma^{-1})$ does not exist (for any d) unless $k - p - 3 > 0$. We assume that the latter inequality holds.

A troublesome result is given by

THEOREM 3.1. *Let $\hat{\Sigma}_a^{-1}$ be given by* (3.1), $0 < d \leq 4$; *and let $\hat{\Sigma}_{a_0}^{-1}$ be the unbiased estimator. If $k > p + 3 > 4$, then*

(a) $R_1(\hat{\Sigma}_{a_0}^{-1}, \Sigma^{-1}) \leq R_1(\hat{\Sigma}_a^{-1}, \Sigma^{-1})$ *for all* Σ, *but*

(b) $R_2(\hat{\Sigma}_{a_0}^{-1}, \Sigma^{-1}) \geq R_2(\hat{\Sigma}_a^{-1}, \Sigma^{-1})$ *for all* Σ.

PROOF. Obviously, $k - p - 3 > 0$; also, it is easy to see that the conditions of Lemma 2.1 are satisfied. From Theorem 2.1 (a), if $d \neq 0$, then $\alpha_1(\Sigma) = d^2E \operatorname{tr} \mathbf{S}^{-1}$, so $\hat{\Sigma}_{a_0}^{-1}$ dominates $\hat{\Sigma}^{-1}$ (mod L_1). From Theorem 2.1 (b), $d > 0$, $\hat{\alpha}_2(\mathbf{S}) = -d(2 - d)\operatorname{tr}(\mathbf{S}^{-2}Q) - 2d(\operatorname{tr} \mathbf{S}^{-1})(\operatorname{tr} \mathbf{S}^{-1}Q)$

$$\leq (d^2 - 4d)\operatorname{tr}(\mathbf{S}^{-2}Q) \tag{3.3}$$

(since $\operatorname{tr}(\mathbf{S}^{-2}Q) \leq (\operatorname{tr} \mathbf{S}^{-1})(\operatorname{tr} \mathbf{S}^{-1}Q)$). Thus, if $0 < d \leq 4$, then $\alpha_2(\Sigma) \leq 0$ ($\forall\Sigma$) and $\hat{\Sigma}^{-1}$ dominates $\hat{\Sigma}_{a_0}^{-1}$(mod L_2). □

Theorem 3.1 is a reversal similar to that described in [4] for the estimators $a\mathbf{S}^{-1} + (b/\operatorname{tr} \mathbf{S})I$, $a = k - p - 1$. In part (b) (Theorem 3.1), note the troublesome cases associated with $d = 4$. If $a_0 = 3$, then $3\mathbf{S}^{-1}$ is dominated by $-\mathbf{S}^{-1}$; i.e., the usual estimator is dominated by a negative definite matrix! Also, if $a_0 = 4$, then $4\mathbf{S}^{-1}$ is dominated by $0_{p\times p}$. These results and those which follow are illustrated in Section 7 by Monte Carlo methods.

4. The estimators $a(\mathbf{S})\mathbf{S}^{-1}$. A generalization of (3.1) is

$$\hat{\Sigma}^{-1} = a_0[1 - t(\cdot)]\mathbf{S}^{-1}, \qquad a_0 = k - p - 1, \tag{4.1}$$

in which $t(\cdot)$ is a function of

$$(4.2)\qquad (\text{tr}\, \mathbf{S})/p, \qquad |\mathbf{S}|^{1/p} \qquad \text{or} \quad p/(\text{tr}\, \mathbf{S}^{-1})$$

(the arithmetic, geometric or harmonic mean eigenvalue of S).

Attention is restricted to loss function L_2, so we require $\hat{\alpha}_2(\mathbf{S}) \leqslant 0$; i.e.,

$$(4.3)\quad (f^2 + 2f)\text{tr}(\mathbf{S}^{-2}Q) + 2f(\text{tr}\, \mathbf{S}^{-1})(\text{tr}\, \mathbf{S}^{-1}Q) - 4\,\text{tr}\left(\frac{\partial f}{\partial S}\right)\cdot T_{(\frac{1}{2})} \leqslant 0,\ T = \mathbf{S}^{-1}Q$$

(see Theorem 2.1 (b)). For the present case, $f(\mathbf{S}) = -a_0 t(\cdot)$.

THEOREM 4.1. *Let $\hat{\Sigma}^{-1}$ be given by* (4.1), *$t(\cdot)$ a nonincreasing function of any argument* (4.2). *Also, let $\hat{\Sigma}_{a_0}^{-1} = a_0 S^{-1}$ where $a_0 = k - p - 1$. If $0 \leqslant t(\cdot) \leqslant 4/a$, then*

$$R_2(\hat{\Sigma}^{-1}, \Sigma^{-1}) \leqslant R_2(\hat{\Sigma}_{a_0}^{-1}, \Sigma^{-1}) \qquad (\forall \Sigma).$$

PROOF. If $f(\mathbf{S}) = -at(\text{tr}\, \mathbf{S}/p)$, then $\partial f/\partial S = -(at'/p)I_{p\times p}$. The left side of (4.3) is $(a^2t^2 - 2at)\text{tr}(\mathbf{S}^{-2}Q) - 2at(\text{tr}\, \mathbf{S}^{-1})(\text{tr}\, \mathbf{S}^{-1}Q) + (4at'/p)\text{tr}(\mathbf{S}^{-1}Q) \leqslant (a^2t^2 - 4at)\text{tr}(\mathbf{S}^{-2}Q) + (4at'/p)\text{tr}(S^{-1}Q) \leqslant 0$ (because $t' \leqslant 0$ and $0 \leqslant t \leqslant 4/a$).

Now, if $f(\mathbf{S}) = -at(|\mathbf{S}|^{1/p})$, then $\partial f/\partial S = -(a|\mathbf{S}|^{1/p}t'/p)\mathbf{S}_{(2)}^{-1}$—see Rao [8], page 72. The left side of (4.3) is

$$(4.4)\quad (a^2t^2 - 2at)\text{tr}(S^{-2}Q) - 2at(\text{tr}\, S^{-1})(\text{tr}\, S^{-1}Q) + \left(\frac{4a|\mathbf{S}|^{1/p}t'}{p}\right)\text{tr}(\mathbf{S}^{-2}Q) \leqslant \left[a^2t^2 - 4at + \frac{4a|\mathbf{S}|^{1/p}t'}{p}\right]\text{tr}(\mathbf{S}^{-2}Q) \leqslant 0$$

(as in the previous case).

Finally, $f(\mathbf{S}) = -at(p/\text{tr}\, \mathbf{S}^{-1})$. From (2.3), $\partial f/\partial S = -ap(\text{tr}\, \mathbf{S}^{-1})^{-2}t' \cdot \mathbf{S}_{(2)}^{-2}$. Our sufficient condition for (4.3) becomes $(a^2t^2 - 4at)\text{tr}(\mathbf{S}^{-2}Q) + (4a/p)(p/\text{tr}\, \mathbf{S}^{-1})^2 \cdot t' \cdot \text{tr}(\mathbf{S}^{-3}Q) \leqslant 0$, which, again, is satisfied by hypothesis. Therefore, if $t(\cdot)$ is a function of any argument (4.2), then $\alpha_2(\Sigma) \leqslant 0$ for all Σ. □

Note that Theorem 4.1 reduces to Theorem 3.1 (b) if $t(\cdot) \equiv d/a_0$. Again, for the case $a_0 = 3$, we can select a "good estimator" which is negative definite on a set of positive probability (let $t(\cdot)$ be positive, nonincreasing and let $t(0) = \frac{4}{3}$, say).

The case $f(\mathbf{S}) = -at(|\mathbf{S}|^{1/p})$ deserves further comment. Denote $|\mathbf{S}|^{1/p}$ by z. From (4.4), it is clear that $\alpha_2(\Sigma) \leqslant 0$ $(\forall\Sigma)$ if

$$(4.5)\qquad at^2(z) - 4t(z) + 4z\, t'(z)\, p \leqslant 0.$$

Thus, the restriction of Theorem 4.1 to nonincreasing $t(\cdot)$ is unnecessary in this case. In fact, the inequality (4.5) has a family of increasing solutions,

$$(4.6)\qquad t(z) = \frac{c_1}{c_2 + (a/z)^p} = \frac{c_1}{c_2 + |a\mathbf{S}^{-1}|}$$

for constants c_i, $i = 1, 2$ (see Section 5). Corresponding to (4.6) is the estimator of Σ^{-1} given by

$$\hat{\Sigma}^{-1} = \left[1 - \frac{c_1}{c_2 + |a\mathbf{S}^{-1}|}\right](a\mathbf{S}^{-1}). \tag{4.7}$$

This, (4.7), is an analogue of Stein's estimator [6] for the mean vector of a multivariate normal distribution.

5. The differential inequalities. Bernoulli's equation is given by

$$t'(x) + A(x)t(x) + B(x)t^2(x) = 0, \tag{5.1}$$

for specified $A(x)$, $B(x) \not\equiv 0$. Throughout, we need solutions to the related inequality,

$$t'(x) + A(x)t(x) + B(x)t^2(x) \leqslant 0. \tag{5.2}$$

In the present discussion, the argument $x \geqslant 0$ is a generic function of S. Note that (4.5) is a special case of (5.2). The equation (5.1) can always be reduced to a linear one, and its solutions provide information about (5.2).

This section contains three lemmas. Lemmas 5.1 and 5.2 give rather obvious sufficient conditions for (5.2); Lemma 5.3 treats the inequality (4.5). From these lemmas, we obtain (4.7) and certain of its generalizations.

We remark that (4.5) also plays an important role in Section 6. There, however, the argument measures eigenvalue dispersion rather than central tendency.

LEMMA 5.1. *Let $t(x)$ and $t_c(x)$ be absolutely continuous functions where $t_c(x)$ is a solution of* (5.1), *c a constant of integration. Let x_1 be arbitrary. If there exists $c = c(x_1)$ such that*

(i) $$t(x_1) = t_c(x_1)$$

and

(ii) $$t'(x_1) \leqslant t_c'(x_1),$$

then $t(x)$ satisfies (5.2).

PROOF. Omitted. ∎

As an equation, (4.5) has the family of solutions

$$t_c(x) = 4/(a + cx^{-p}), \qquad c \geqslant 0. \tag{5.3}$$

If $0 < t(x) \leqslant 4/a$, we can always find c as required in (i) and (ii) above.

Let $t_c(x)$ be given by (5.3), and let

$$t(x) = \delta t_c(x) = (4\delta)/(a + cx^{-p}), \qquad 0 < \delta < 1. \tag{5.4}$$

(This function was stated in (4.6).) Obviously, $t(x)$ satisfies (4.5). In general, we have

LEMMA 5.2. *Given $0 < \delta < 1$, if $t_c(x)$ satisfies* (5.1), *then $t(x) = \delta t_c(x)$ satisfies* (5.2) *with strict inequality.*

PROOF. Omitted. ∎

Lemma 5.3 is motivated by a result of Efron and Morris [2]. The lemma shows that we can generalize (4.7) in the same manner that they [2] generalized the James-Stein estimator of a mean vector.

LEMMA 5.3. *Let* $\psi(x) = x^{-p}t(x)/[(4/a) - t(x)]$. *The inequality* (4.5) *holds* iff $\psi'(x) < 0$.

PROOF. A simple calculation shows that $\psi'(z) < 0$ is equivalent to (4.5). ▯

6. Random mixtures of S^{-1} and I. Assume that

$$\hat{\Sigma}^{-1} = a[1 - t_1(U)]S^{-1} + [bt_2(U)/\text{tr } S]I, \tag{6.1}$$

$0 < a \leq k - p - 1$, $b > 0$, $t_j(U)$ nondecreasing, $0 \leq t_j(U) \leq 1$, $j = 1, 2$, with U given by

$$U_1 = p|S|^{1/p}/\text{tr } S \quad \text{or} \quad U_2 = p^2/[(\text{tr } S)(\text{tr } S^{-1})]. \tag{6.2}$$

For $t_1(U_1) \equiv 0$, Haff [4] gave conditions under which $\hat{\Sigma}^{-1}$ dominates $\hat{\Sigma}_a^{-1}$ (mod L_i), $i = 1, 2$. We do not have such results for loss function L_1, $t_1(U) \not\equiv 0$. The results which follow are mostly L_2 results. We note, however, that Efron and Morris [2] studied the special case $t_1(U) = t_2(U) = \gamma$ (a constant) under L_1, $0 < \gamma < 1$.

For the estimator (6.1), $f(S) = -at_1(U)$ and $g(S) = bt_2(U)/\text{tr } S$. We need $\partial f/\partial S$ and $\partial g/\partial S$—see Theorem 2.1. These derivatives are as follows:

If $U = U_1$, then

$$\begin{aligned} \partial f/\partial S &= aUt_1'(U)[(1/\text{tr } S)I_{p\times p} - (1/p)S_{(2)}^{-1}], \quad \text{and} \\ \partial g/\partial S &= -b(1/\text{tr } S)^2[t_2(U) + Ut_2'(U)]I_{p\times p} \\ &\quad + (b/p \text{ tr } S)Ut_2'(U)S_{(2)}^{-1}. \end{aligned} \tag{6.3}$$

If $U = U_2$, then

$$\partial f/\partial S = -aUt_1'(U)[(1/\text{tr } S^{-1})S_{(2)}^{-2} - (1/\text{tr } S)I], \quad \text{and} \tag{6.4}$$

$$\partial g/\partial S = bt_2'(U)[(U/p)^2S_{(2)}^{-2} - U/(\text{tr } S)^2 I] - bt_2(U)/(\text{tr } S)^2 \cdot I.$$

The derivatives (6.3) were given by Haff [4]. Those of (6.4) follow from (2.3).

Our first specialization of Theorem 3.1 is

COROLLARY 6.1. *For* $U = U_1$, *we have* $\alpha_1(\Sigma) \leq 0$ $(\forall\Sigma)$ *if*

$$\begin{aligned} &(pat_1)^2 + 4apU_1(1 - U_2)t_1' \\ &\quad + U_2[4bt_2 - 2abpt_1t_2 - 2bp(p + 1)t_2 + b^2t_2^2] \leq 0 \end{aligned} \tag{6.5}$$

for all $\mathbf{S}$; *and* $\alpha_2(\Sigma) \leqslant 0$ $(\forall\Sigma)$ *if*

$$\begin{aligned}
&\left[a^2t_1^2 - 2a(a^* + 1)t_1 + 4aU_1(1/p)t_1'\right]\mathrm{tr}(S^{-2}Q)(\mathrm{tr}\,\mathbf{S})^2/\mathrm{tr}\,Q \\
&+\left[2a^*bt_2 - 2abt_1t_2 - 4bU_1(1/p)t_2' - 4aU_1t_1'\right]\mathrm{tr}(S^{-1}Q)(\mathrm{tr}\,\mathbf{S})/\mathrm{tr}\,Q \\
&- 2at_1(\mathrm{tr}\,\mathbf{S}^{-1})(\mathrm{tr}\,\mathbf{S}^{-1}Q)(\mathrm{tr}\,\mathbf{S})^2/\mathrm{tr}\,Q \\
&+\left[4bt_2 + 4bU_1t_2' + (bt_2)^2\right] \leqslant 0, \qquad \text{for all } \mathbf{S}.
\end{aligned} \tag{6.6}$$

PROOF. Straightforward. □

We note some complications for loss function L_1. For $t_1(U_1) \equiv 0$, we obtain $\alpha_1(\Sigma) < 0$ $(\forall\Sigma)$ if $0 \leqslant t_2(U_1) \leqslant (2/b)(p^2 + p - 1)$. The latter is a result from [4]. However, if $t_1(U_1) \neq 0$, then (6.5) depends on U_1 and U_2, and it is doubtful that $\alpha_1(\Sigma) < 0$ $(\forall\Sigma)$. If $U = U_2$, $t_1(U_2) \not\equiv 0$, then the counterpart of (6.5) can be reduced to an inequality in U_2 only; still, we are unable to establish dominance (mod L_1).

For loss function L_2, the following theorem shows that we can strictly improve upon the choice $t_1(U_1) \equiv 0$.

THEOREM 6.1. *Let* $U = U_1$, $a = k - p - 1(a^* = 0)$, $b > 0$, *and* $g(\mathbf{S}) = bt_2(U)/\mathrm{tr}\,\mathbf{S}$, $t_2(U)$ *nondecreasing*, $0 \leqslant t_2(U) \leqslant 1$. *Also, let* $\hat{\Sigma}^{-1}{}_* = a\mathbf{S}^{-1} + [bt_2(U)/\mathrm{tr}\,\mathbf{S}]I$ *and* $\hat{\Sigma}^{-1} = a[1 - t_1(U)]\mathbf{S}^{-1} + [bt_2(U)/\mathrm{tr}\,\mathbf{S}]I$ *in which* $t_1(U)$ *is a nondecreasing solution of*

$$4Ut'(U) - 4pt(U) + apt^2(U) \leqslant 0. \tag{6.7}$$

Then $\hat{\Sigma}^{-1}$ *dominates* $\hat{\Sigma}_*^{-1}$(mod L_2).

PROOF. From (6.6), it is sufficient to show that $[a^2t_1^2 - 2at_1 + 4aU(t_1'/p)]\mathrm{tr}(\mathbf{S}^{-2}Q) - 2at_1[\mathrm{tr}\,\mathbf{S}^{-1}][\mathrm{tr}\,\mathbf{S}^{-1}Q] \leqslant 0$. Since $[\mathrm{tr}\,\mathbf{S}^{-1}][\mathrm{tr}\,\mathbf{S}^{-1}Q] \geqslant \mathrm{tr}(\mathbf{S}^{-2}Q)$, the desired inequality holds if $4aU(t_1'/p) - 4at_1 + a^2t_1^2 \leqslant 0$, and the proof is complete. □

In Theorem 6.1, the assumption $a^* = 0$ was for convenience only. A similar result holds for $a^* < 0$. Recall that (6.7) was given (except for the argument) by (4.5), a family of solutions by (4.6).

The remaining results are for random convex mixtures of aS^{-1} and $(b/\mathrm{tr}\,\mathbf{S})I$; i.e.,

$$\hat{\Sigma}_1^{-1} = a[1 - t(U)]\mathbf{S}^{-1} + (bt(U)/\mathrm{tr}\,\mathbf{S})I, \tag{6.8}$$

$(t_j(U) = t(U), j = 1, 2)$. For convenience, let us record

$$\begin{aligned}
&\text{(a)}\quad \mathrm{tr}(S^{-N}Q)/\mathrm{tr}Q \geqslant |\mathbf{S}|^{-N/p}Q^*, \\
&\qquad N \text{ a positive integer, } Q^* \equiv p|Q|^{1/p}/\mathrm{tr}\,Q, \\
&\text{(b)}\quad (\mathrm{tr}\,\mathbf{S}^{-1})(\mathrm{tr}\,\mathbf{S}^{-N}Q) \geqslant \mathrm{tr}(\mathbf{S}^{-(N+1)}Q), \\
&\text{(c)}\quad \mathrm{tr}\,\mathbf{S}^{-2}/(\mathrm{tr}\,\mathbf{S}^{-1})^2 \geqslant 1/p.
\end{aligned} \tag{6.9}$$

In (6.9), (a) is apparent from [1], Exercise 4, page 134; (b) follows from the spectral decomposition of S; and (c) is from [1], Exercise 13, page 137.

THEOREM 6.2. *Assume that* $\hat{\Sigma}^{-1}$ *is given by* (6.8) *with*
(i) $U = U_1$,
(ii) $t(U) \geq 0$ *a nondecreasing solution of* $at^2 - 4t + 4Ut'/p \leq 0$,
(iii) $a = k - p - 1$ ($a^* = 0$), $b = ap$,
(iv) $pQ^* > 1$.
Then $\hat{\Sigma}^{-1}$ *dominates* $\hat{\Sigma}_{a_0}^{-1} = (k - p - 1)\mathbf{S}^{-1}(\text{mod } L_2)$.

PROOF. If we apply (6.9)(c) to the second line following (6.6), we obtain $\alpha_2(\Sigma) \leq 0\ (\forall\Sigma)$ if

$$\begin{aligned} &\left[a^2t^2 - 4at + 4aUt'/p\right]\operatorname{tr}(\mathbf{S}^{-2}Q)(\operatorname{tr}\mathbf{S})^2/\operatorname{tr}Q \\ &-\left[2abt^2 + 4bUt'/p + 4aUt'\right]\operatorname{tr}(\mathbf{S}^{-1}Q)(\operatorname{tr}\mathbf{S})/\operatorname{tr}Q \\ &+\left[4bt + 4bUt' + b^2t^2\right] \leq 0 \qquad (\forall \mathbf{S}). \end{aligned} \tag{6.10}$$

The coefficients of $\operatorname{tr}(\mathbf{S}^{-2}Q)(\operatorname{tr}\mathbf{S})^2/\operatorname{tr}Q$ and $\operatorname{tr}(\mathbf{S}^{-1}Q)(\operatorname{tr}\mathbf{S})/\operatorname{tr}Q$ are negative by hypothesis, so from (6.9) (a), a sufficient condition for (6.10) is $[a^2t^2 - 4at + 4aUt'/p](p/U)^2Q^* - [2abt^2 + 4bUt'/p + 4aUt'](p/U)Q^* + [4bt + 4bUt' + b^2t^2] \leq 0$. Since $0 < U \leq 1$ (a.e.), a sufficient condition for the latter is $[a^2t^2 - 4at + 4aUt'/p]p^2Q^* - [2abt^2 + 4bUt'/p + 4aUt']pQ^* + [4bt + 4bUt' + b^2t^2] \leq 0$, or, after simplification, $(ap - b)^2t^2Q^* + (4bUt' - 4ap^2t + b^2t^2)(1 - Q^*) + 4bt - 4ap^2tQ^* \leq 0$. After substituting $b = ap$, we obtain $ap^2(at^2 - 4t + 4Ut'/p)(1 - Q^*) + 4apt(1 - pQ^*) \leq 0$, which is true by hypothesis. The proof is complete. □

Finally, let U be given by U_2, $\partial f/\partial S$ and $\partial g/\partial S$ by (6.4). Here we assume $Q = I (Q^* = 1)$, so that $L_2(\hat{\Sigma}^{-1}, \Sigma^{-1}) = \Sigma\Sigma(\hat{\sigma}^{ij} - \sigma^{ij})^2$. From Theorem 2.1(b), $a^* = 0$, we obtain $\alpha_2(\Sigma) \leq 0\ (\forall\Sigma)$ if

$$\begin{aligned} &\left[4aUt'\right]\operatorname{tr}(\mathbf{S}^{-3})/(\operatorname{tr}\mathbf{S}^{-1})^3 \\ &+\left[a^2t^2 - 2at - 4b(U/p)^2t'\right]\operatorname{tr}(\mathbf{S}^{-2})/(\operatorname{tr}\mathbf{S}^{-1})^2 \\ &+\left[-2abt^2 - 4aUt'\right]/\left[(\operatorname{tr}\mathbf{S}^{-1})(\operatorname{tr}\mathbf{S})\right] \\ &+\left[4bUt' + 4bt + b^2t^2\right]p/\left[(\operatorname{tr}\mathbf{S})(\operatorname{tr}\mathbf{S}^{-1})\right]^2 - 2\,at \leq 0. \end{aligned} \tag{6.11}$$

In (6.11), note that

$$\begin{aligned} \operatorname{tr}\mathbf{S}^{-3}/(\operatorname{tr}\mathbf{S}^{-1})^3 &\leq (\operatorname{tr}\mathbf{S}^{-1})(\operatorname{tr}\mathbf{S}^{-2}Q)/(\operatorname{tr}\mathbf{S}^{-1})^3 \\ &= \operatorname{tr}(\mathbf{S}^{-2}Q)/(\operatorname{tr}\mathbf{S}^{-1})^2 \\ &\leq 1. \end{aligned} \tag{6.12}$$

THEOREM 6.3. *Assume* $\hat{\Sigma}^{-1}$ *is given by* (6.8) *with*

(i) $U = U_2$,

(ii) $t(U)$ *a nondecreasing solution of* $at^2 - 4t + 4Ut'/p \leq 0$ and $2Ut' \leq t$,

(iii) $a = k - p - 1$ $(a^* = 0)$, $b = ap$.

Then $\hat{\Sigma}^{-1}$ *dominates* $\hat{\Sigma}_{a_0}^{-1} = (k - p - 1)S^{-1}$ (mod L_2).

PROOF. From (6.12) and simple algebra, we obtain the following sufficient condition for (6.11):

$$\begin{aligned}
&\left[(1 - 1/p)4aUt'\right]\operatorname{tr} \mathbf{S}^{-2}/(\operatorname{tr} \mathbf{S}^{-1})^2 \\
&+\left[a^2t^2 - 2at + \frac{4aUt'}{p}\right]\left[\operatorname{tr} \mathbf{S}^{-2}/(\operatorname{tr} \mathbf{S}^{-1})^2\right] - 2at \\
&+\left[-2abt^2 - 4aUt'\right]/\left[(\operatorname{tr} \mathbf{S})(\operatorname{tr} \mathbf{S}^{-1})\right] \\
&-\left[4b(U/p)^2t'\right]\operatorname{tr} \mathbf{S}^{-2}/(\operatorname{tr} \mathbf{S}^{-1})^2 \\
&+\left[4bUt' + 4bt + b^2t^2\right]p/\left[(\operatorname{tr} \mathbf{S})(\operatorname{tr} \mathbf{S}^{-1})\right]^2 \leq 0.
\end{aligned} \tag{6.13}$$

Recall that $U = p^2/[(\operatorname{tr} \mathbf{S})(\operatorname{tr} \mathbf{S}^{-1})]$. In the first line of (6.13) replace $\operatorname{tr} \mathbf{S}^{-2}/(\operatorname{tr} \mathbf{S}^{-1})^2$ by 1; in the second and fourth lines, replace it by $1/p$ (recall (6.9)(c)). This done, we have the sufficient condition

$$\begin{aligned}
&(p^3/U^2)(1 - 1/p)(4aUt' - 2at) \\
&+ \left(a^2t^2 - 4at + \frac{4aUt'}{p}\right)(p/U)^2 \\
&- \left(2abt^2 + 4aUt' + \frac{4bUt'}{p}\right)(p/U) \\
&+ (4bUt' + 4bt + b^2t^2) \leq 0.
\end{aligned} \tag{6.14}$$

The first term in (6.14) is negative by hypothesis; and the sum of the others is negative from Theorem 6.2—see (6.11) and following. The proof is complete. □

Let $t(\cdot)$ be given by (5.4). (Recall that $t(U)$ satisfies the first inequality in Theorem 6.3 (ii)). An easy calculation shows that $t(U)$ also satisfies $2Ut' \leq t$ for sufficiently small c.

7. Simulation results. Consider the special case $\mathbf{S}_{3\times 3} \sim W(\Sigma, 7)$, i.e., $k = 7$, $p = 3$. Given below are the results of a computer simulation in which we compared

$\hat{\Sigma}_3^{-1} = 3\mathbf{S}^{-1}$, the unbiased estimator;

$\hat{\Sigma}_{-1}^{-1} = -\mathbf{S}^{-1}$, a negative definite estimator!;

$\hat{\Sigma}^{-1} = 3[1 - U^{1/2}]\mathbf{S}^{-1} + \left(\dfrac{9U^{1/2}}{\operatorname{tr} \mathbf{S}}\right)I$, $U = 3|\mathbf{S}|^{1/3}/\operatorname{tr} \mathbf{S}$, a special case of (6.1);

and

$\hat{\Sigma}_S^{-1} = [1 - 3|\mathbf{S}|/(1 + 3|\mathbf{S}|)](3\mathbf{S}^{-1})$, a special case of (4.7). The loss function was L_2, $Q = I$; i.e.,

$$L_2(\hat{\Sigma}^{-1}, \Sigma^{-1}) = \Sigma\Sigma(\hat{\sigma}^{ij} - \sigma^{ij})^2. \tag{7.1}$$

Denote the eigenvalues of Σ by $\lambda_1, \lambda_2, \lambda_3$. We specified five diagonal matrices,

$$D = \text{diag}(\lambda_1, \lambda_2, \lambda_3),$$

and for each, the following was replicated 1000 times:

$$\text{Let } \phi \sim \text{Uniform } [0, \Pi],$$
$$\theta \sim \text{Uniform } [0, 2\Pi].$$

From a realization of (ϕ, θ), we computed the orthogonal matrix

$$R = \begin{bmatrix} \sin\phi \quad \cos\theta & \cos\phi \quad \cos\theta & -\sin\theta \\ \sin\phi \quad \sin\theta & \cos\phi \quad \sin\theta & \cos\theta \\ \cos\phi & -\sin\phi & 0 \end{bmatrix},$$

and defined $\Sigma^{-1} = RD^{-1}R^t$ as the matrix to be estimated. Then, independent vector observations $X_{(1)}, X_{(2)}, \cdots, X_{(7)}$ were obtained from a $N_3(0, I)$ distribution, and the transforms $Y_{(i)} = RD^{\frac{1}{2}}X_{(i)}$ were computed, $i = 1, \cdots, 7$. From the Y's, which are i.i.d. $N_3(0, \Sigma)$, we computed

$$S_{3\times 3} = \Sigma_{i=1}^{7} Y_{(i)} Y_{(i)}^t \sim \text{Wishart } (\Sigma, 7).$$

Finally, for each of $\hat{\Sigma}_3^{-1}$, $\hat{\Sigma}_{-1}^{-1}$, etc., we computed (7.1). The following table contains means of (7.1), for each estimator, and estimated standard deviations of means (over $N = 1000$ orientations).

In Table 1, note that $\hat{\Sigma}_{-1}^{-1}$ "outperformed" $\hat{\Sigma}_3^{-1}$ at each set of eigenvalues except $\{1, 5.5, 10\}$. Note, also, that $\hat{\Sigma}^{-1}$ and $\hat{\Sigma}_S^{-1}$ significantly outperformed $\hat{\Sigma}_3^{-1}$. Other simulations (unpublished) indicate that the latter comparisons are more pronounced still for larger values of p.

TABLE 1[1]
Estimation of quadratic risk (R_2) for each of four estimators.

Specified eigenvalues			Mean loss and estimated standard deviation of the mean (N = 1000 orientations)			
λ_1	λ_2	λ_3	$\hat{\Sigma}_{-1}^{-1}$	$\underline{\underline{\hat{\Sigma}_3^{-1}}}$	$\hat{\Sigma}^{-1}$	$\hat{\Sigma}_S^{-1}$
.9	.5	.1	205.62	305.06	75.64	108.88
			(7.43)	(91.04)	(14.58)	(15.99)
1	1	1	6.07	6.50	1.92	2.94
			(0.17)	(1.03)	(0.59)	(0.01)
1	2.5	4	2.49	2.62	1.15	1.21
			(0.09)	(0.55)	(0.40)	(0.01)
1	4	7	2.21	2.39	0.89	1.08
			(0.10)	(0.60)	(0.21)	(0.01)
1	5.5	10	2.95	2.18	0.61	1.04
			(1.01)	(0.53)	(0.02)	(0.01)

[1]The first entry is mean loss; the second (in parentheses) is estimated standard deviation of the mean.

Clearly $\hat{\Sigma}_3^{-1}$ is a bad estimator, but it cannot be *this* bad. The loss function L_2 must be suspect—especially when $k - p$ is small. Perhaps a more respectable loss function is

$$L(\hat{\Sigma}^{-1}, \Sigma^{-1}) = \text{tr}(\hat{\Sigma}^{-1}\Sigma - I)^2 \tag{7.2}$$

which is invariant under nonsingular linear transformations of the original data vectors. The latter, (7.2), is treated by the author in a forthcoming paper [5].

Acknowledgment. I would like to thank the referee and Associate Editor for many useful comments and suggestions.

REFERENCES

[1] BELLMAN, R. (1970). *Introduction to Matrix Analysis*. McGraw-Hill, New York.

[2] EFRON, B. and MORRIS, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **4** 11–21.

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

[4] HAFF, L. R. (1977). Minimax estimators for a multinormal precision matrix. *J. Multivariate Anal.* **7** 374–385.

[5] HAFF, L. R. (1978). A fundamental identity for the Wishart distribution with applications. Unpublished manuscript.

[6] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probability* **1** 361–379. Univ. California Press.

[7] PRESS, S. J. (1972). *Applied Multivariate Analysis*. Holt, Rinehart and Winston, New York.

[8] RAO, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley, New York.

[9] STEIN, C. (1973). Estimation of the mean of a multivariate normal distribution, *Proc. Prague Symp. Asymptotic Statist.* 345–381.

[10] STEIN, C., EFRON, B., and MORRIS, C (1972). Improving the usual estimator of a normal covariance matrix. Technical Report No. 37, Depart. Statist., Stanford Univ.

DEPARTMENT OF MATHEMATICS, C-012,
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093