The Annals of Statistics
1985, Vol. 13, No. 4, 1581–1591

ESTIMATION OF A COVARIANCE MATRIX UNDER STEIN'S LOSS

BY DIPAK K. DEY[1] AND C. SRINIVASAN[2]

Texas Tech University and University of Kentucky

Stein's general technique for improving upon the best invariant unbiased and minimax estimators of the normal covariance matrix is described. The technique is to obtain solutions to a certain differential inequality involving the eigenvalues of the sample covariance matrix. Several improved estimators are obtained by solving the differential inequality. These estimators shrink or expand the sample eigenvalues depending on their magnitude. A scale invariant, adaptive minimax estimator is also obtained.

1. Introduction. In this paper we consider the problem of estimating the covariance matrix Σ, of a multivariate normal population. The usual estimator is S/k, where S is distributed according to the Wishart distribution $W_p(\Sigma, k)$. James and Stein (1961) obtained a minimax estimator by considering the best invariant estimator with respect to the triangular group G_T^+ (the group consisting of lower triangular matrices with positive diagonal elements). This estimator, of course, depends on the coordinate system [see also Selliah (1964), Olkin and Selliah (1977)]. Later on, Takemura (1984) obtained an improved orthogonally invariant minimax estimator by averaging Stein's minimax estimator over the $p \times p$ orthogonal matrices with respect to Haar measure. For higher dimensions, however, his estimator does not have a simple form.

Although S/k is unbiased it is known that the sample eigenvalues of S tend to be more spread out than the population eigenvalues of Σ. This fact suggests that one should shrink or expand the sample eigenvalues depending on their magnitude. The works along this direction can be found in Stein (1975, 1977a, b), Efron and Morris (1976), Haff (1977, 1979a, b, 1982), Eaton (1970), and a host of others.

In this paper we consider improved estimators of Σ under a scale invariant loss function introduced in James and Stein (1961). Suppose $\hat{\Sigma}$ is an estimator of Σ that is assumed to incur a loss

$$L(\hat{\Sigma}, \Sigma) = \operatorname{tr}(\hat{\Sigma}\Sigma^{-1}) - \log\det(\hat{\Sigma}\Sigma^{-1}) - p. \tag{1.1}$$

Suppose, further, that an estimator's performance is evaluated by considering the risk function

$$R(\hat{\Sigma}, \Sigma) = E\big(L(\hat{\Sigma}, \Sigma) \mid \Sigma\big). \tag{1.2}$$

The constant risk minimax estimator proposed by James and Stein (1961) is of the form

$$\hat{\Sigma}^M = TDT^t, \tag{1.3}$$

Received February 1984; revised April 1985.

[1] Now at the University of Connecticut, Storrs, CT 06268.

[2] Research supported by NSF Grant number MCS 8212968.

AMS 1980 *subject classifications*. Primary 62F10; secondary 62C99.

Key words and phrases. Covariance matrix, Wishart distribution, Stein's loss, orthogonally invariant estimators, minimax estimators.

where $D = \text{diag}(d_1, d_2, \ldots, d_p)$, $T \in G_T^+$ with $TT^t = S$ and

$$d_i = 1/(k+p+1-2i), \qquad i = 1, 2, \ldots, p. \tag{1.4}$$

With this choice of d_i, the minimax risk is obtained as

$$R(\hat{\Sigma}^M, \Sigma) = \sum_{i=1}^{k} \left[\log(k+p+1-2i) - E(\log \chi^2_{k-i+1})\right], \tag{1.5}$$

where χ^2_{k-i+1} denotes a chi-square random variable with $k-i+1$ degrees of freedom. This is uniformly smaller than the risk of $\hat{\Sigma}_0$. Later Stein (1975, 1977a, b) considered the class of orthogonally invariant estimators, i.e., those of the form

$$\hat{\Sigma} = R\varphi(L)R^t, \tag{1.6}$$

where $S = RLR^t$ with R the matrix of normalized eigenvectors ($RR^t = R^tR = I$), $L = \text{diag}(l_1, l_2, \ldots, l_p)$ is the diagonal matrix of corresponding eigenvalues with $l_1 \geq l_2 \geq \cdots \geq l_p$, and $\varphi(L) = \text{diag}(\varphi_1(L), \varphi_2(L), \ldots, \varphi_p(L))$. Thus $\varphi^{(0)}(L) = (1/k)L$ gives the best invariant and unbiased estimator $\hat{\Sigma}_0 = S/k$ under (1.1). For this loss function, Stein (1975) proposed the particular estimator determined by

$$\varphi_i^s = l_i \Big/ \left[(k-p+1) + 2l_i \sum_{j \neq i} 1/(l_i - l_j)\right], \qquad i = 1, 2, \ldots, p, \tag{1.7}$$

which he obtained by an approximate minimization of the unbiased estimator of the risk function. However, in this case it is not generally true that $\varphi_1^s \geq \varphi_2^s \geq \cdots \geq \varphi_p^s \geq 0$. Stein (1975, 1977a, b) gave an isotonized modification of his estimator [cf., Lin (1977), Lin and Perlman (1985)]. The risk function of Stein's estimator is very complicated, and it has not been determined that Stein's estimator dominates $\hat{\Sigma}_0$. However the Monte Carlo simulation results of Lin and Perlman (1985) indicate that Stein's modified estimator outperforms not only $\hat{\Sigma}_0$ but also the minimax estimator (1.3) significantly over a wide range of Σ's. Recently Haff (1982) studied the form of Bayes estimators of Σ and observed that a slight modification of Stein's (1975) estimator emerges as an approximation of the Bayes rule.

In this paper, we consider the class of orthogonally invariant estimators (1.6). We apply Stein's technique, whereby an unbiased estimator of the risk of an orthogonally invariant estimator is obtained, involving eigenvalue estimators and their derivatives. The improved estimators are obtained by solving a differential inequality for these eigenvalue estimators. Several solutions are obtained which are motivated from Dey, Ghosh, and Srinivasan (1984). In Section 2, we first illustrate our method by presenting an estimator that has a very simple form and dominates $\hat{\Sigma}_0$ in terms of risk. In Section 3, we consider the more important problem of improving upon the constant risk minimax estimator of James and Stein (1961). One of the orthogonally invariant minimax estimators in Section 3 shrinks or expands the sample eigenvalues about a point that can be chosen adaptively.

In Section 4, we briefly mention simulation results for the risk behavior of the proposed minimax estimators. It is observed that these estimators have uniformly smaller risk than the minimax risk, for those p, k, and Σ chosen in the design.

2. Estimators which dominate $\hat{\Sigma}_0$. We first observe that for an estimator $\hat{\Sigma}$ of the form (1.6)

$$L(\hat{\Sigma}, \Sigma) = \operatorname{tr}(\hat{\Sigma}\Sigma^{-1}) - \sum_{i=1}^{p} \log \varphi_i(L) + \log\det(\Sigma) - p. \tag{2.1}$$

Clearly the last two terms are constant with respect to $\hat{\Sigma}$. Thus we define

$$R^*(\hat{\Sigma}, \Sigma) = E\left[\operatorname{tr}(\hat{\Sigma}\Sigma^{-1}) - \sum_{i=1}^{p} \log \varphi_i(L) | \Sigma\right]. \tag{2.2}$$

From (2.1) and (2.2) it follows that the difference in risk between $\hat{\Sigma}$ and $\hat{\Sigma}_0$ is

$$\begin{aligned} \alpha(\Sigma) &= R^*(\hat{\Sigma}, \Sigma) - R^*(\hat{\Sigma}_0, \Sigma) \\ &= E\left[\operatorname{tr}(\hat{\Sigma}\Sigma^{-1} - \hat{\Sigma}_0\Sigma^{-1}) - \sum_{i=1}^{p} \log \frac{\varphi_i(L)}{\varphi_i^0(L)} \middle| \Sigma\right]. \end{aligned} \tag{2.3}$$

Now we will state the following lemma due to Stein and Haff which is needed to evaluate $\alpha(\Sigma)$. The proof is given in Stein (1977a) and Haff (1982).

LEMMA 2.1. *The unbiased estimator of $R^*(\hat{\Sigma}, \Sigma)$ is given as*

$$\begin{aligned} \hat{R}^*(\hat{\Sigma}, \Sigma) &= 2 \sum_{i=1}^{p} \sum_{t>i} \frac{\varphi_i - \varphi_t}{l_i - l_t} + 2 \sum_{i=1}^{p} \frac{\partial \varphi_i}{\partial l_i} \\ &\quad + (k - p - 1) \sum_{i=1}^{p} \frac{\varphi_i}{l_i} - \sum_{i=1}^{p} \log \varphi_i. \end{aligned} \tag{2.4}$$

From Lemma 2.1, it follows that $\hat{R}^*(\hat{\Sigma}, \Sigma)$ depends only on the sample eigenvalues of S. Let us define $\alpha(L)$ as an unbiased estimator of $\alpha(\Sigma)$ [see (2.9)]. We now need the following two lemmas to obtain an upper bound for $\alpha(L)$.

LEMMA 2.2. *For $|x| \le u < 1$,*

$$\log(1 + x) \ge x - \frac{3 - u}{6(1 - u)} x^2. \tag{2.5}$$

PROOF. From the series expansion of $\log(1 + x)$, it follows that

$$\begin{aligned} \log(1 + x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \\ &\ge x - \frac{x^2}{2} - \frac{|x|^3}{3} - \frac{|x|^4}{4} - \cdots \\ &\ge x - \frac{x^2}{2} - \frac{|x|^3}{3}(1 - |x|) \\ &\ge x - \frac{x^2}{2} - \frac{u x^2}{3(1 - u)} = x - \frac{(3 - u)x^2}{6(1 - u)}, \end{aligned}$$

which completes the proof. □

LEMMA 2.3. *If* $x_1 > x_2 > \cdots > x_p$, *then*

$$(p+1)\sum_{i=1}^{p} x_i \ge 2\sum_{i=1}^{p} i x_i. \tag{2.6}$$

PROOF. The proof follows by induction on p. □

The improved estimator of $\hat{\Sigma}_0$ will be given in the following theorem.

THEOREM 2.1. *Consider an estimator* $\hat{\Sigma}$ *of* Σ *as given in* (1.6). *Then* $\hat{\Sigma}$ *will dominate* $\hat{\Sigma}_0$ *in terms of risk if* $p \ge 3$ *and* $\varphi_i(L)$ *is given as*

$$\varphi_i(L) = \frac{l_i}{k} - \frac{(l_i \log l_i)\tau(u)}{b+u}, \tag{2.7}$$

where $u = \sum_{i=1}^{p} \log^2 l_i$, *and* b *is a constant,* $b > 144(p-2)^2/25k^2$, *and* $\tau(u)$ *is a function satisfying*

$$\text{(i) } 0 < \tau(u) < 2(p-2)/k^*, \qquad k^* = 5k^2/6;$$
$$\text{(ii) } \tau(u) \text{ monotone nondecreasing in } u \text{ and } E[\tau'(u)] < \infty. \tag{2.8}$$

PROOF. Define $\varphi_i(L) = \varphi_i^{(0)}(L) + \gamma_i(L)$, where $\varphi_i^{(0)}(L) = l_i/k$, $i = 1, 2, \dots, p$, and $\gamma_i(L)$ will be appropriately chosen. Using Lemma 2.1, it follows that

$$\begin{aligned}\alpha(L) &= 2\sum_{i=1}^{p}\sum_{t>i} \frac{\gamma_i(L) - \gamma_t(L)}{l_i - l_t} + 2\sum_{i=1}^{p} \frac{\partial \gamma_i(L)}{\partial l_i} \\ &\quad + (k-p-1)\sum_{i=1}^{p} \frac{\gamma_i(L)}{l_i} - \sum_{i=1}^{p} \log\left(1 + \frac{k\gamma_i(L)}{l_i}\right).\end{aligned} \tag{2.9}$$

It is sufficient to find a solution $(\gamma_1(L), \gamma_2(L), \dots, \gamma_p(L))$ to the differential inequality $\alpha(L) \le 0$, with strict inequality for some set of l with positive measure. For notational convenience let us define

$$\gamma_i(L) = l_i \xi(l_i)/B(L), \qquad i = 1, 2, \dots, p, \tag{2.10}$$

where $\xi(l_i) = \log l_i$, $i = 1, 2, \dots, p$, and $B^{-1}(L) = -\tau(u)/(b+u)$. We will show that (2.10) is a solution of $\alpha(L) \le 0$.

Now it is easy to observe that

$$\begin{aligned} k\left|\frac{\gamma_i(L)}{l_i}\right| &= \frac{k\tau(u)}{b+u}|\log l_i| \\ &\le \frac{k\tau(u)}{2\sqrt{b}} < \frac{2k(p-2)}{2\sqrt{b}\,k^*} = \frac{6(p-2)}{5k\sqrt{b}} < \frac{1}{2}. \end{aligned}$$

Hence using Lemma 2.2, with $u = \frac{1}{2}$, one gets

$$\log\left(1 + \frac{k\gamma_i(L)}{l_i}\right) \geq \frac{k\gamma_i(L)}{l_i} - \frac{5}{6}\frac{k^2\gamma_i^2(L)}{l_i^2}.$$

Thus from (2.9) it follows that

$$\begin{aligned}\alpha(L) &\leq 2\sum_{i=1}^{p}\sum_{t>i}\frac{\gamma_i(L)-\gamma_t(L)}{l_i - l_t} + 2\sum_{i=1}^{p}\frac{\partial\gamma_i(L)}{l_i}\\ &\quad +(k-p-1)\sum_{i=1}^{p}\frac{\gamma_i(L)}{l_i} - k\sum_{i=1}^{p}\frac{\gamma_i(L)}{l_i} + \frac{5k^2}{6}\sum_{i=1}^{p}\frac{\gamma_i^2(L)}{l_i^2}\\ &= 2\sum_{i=1}^{p}\sum_{t>i}\frac{\gamma_i(L)-\gamma_t(L)}{l_i-l_t} - (p+1)\sum_{i=1}^{p}\frac{\gamma_i(L)}{l_i}\\ &\quad +2\sum_{i=1}^{p}\frac{\partial\gamma_i(L)}{\partial l_i} + \frac{5k^2}{6}\sum_{i=1}^{p}\frac{\gamma_i^2(L)}{l_i^2}.\end{aligned}$$

Now using (2.10) it follows that

$$\begin{aligned}\alpha(L) &\leq \frac{2}{B(L)}\sum_{i=1}^{p}\sum_{t>i}\frac{l_i\xi(l_i)-l_t\xi(l_t)}{l_i-l_t} - \frac{(p+1)}{B(L)}\sum_{i=1}^{p}\xi(l_i)\\ &\quad +2\sum_{i=1}^{p}\left\{\frac{\xi(l_i)}{B(L)} + l_i\frac{\partial}{\partial l_i}\left(\frac{\xi(l_i)}{B(L)}\right)\right\} + \frac{5k^2}{6}\sum_{i=1}^{p}\frac{\xi^2(l_i)}{B^2(L)}\\ &= \frac{2}{B(L)}\sum_{i=1}^{p}\sum_{t>i}\frac{l_i\xi(l_i)-l_t\xi(l_t)+l_t(\xi(l_i)-\xi(l_t))}{l_i-l_t}\\ &\quad -\frac{(p-1)}{B(L)}\sum_{i=1}^{p}\xi(l_i) + 2\sum_{i=1}^{p}l_i\frac{\partial}{\partial l_i}\left(\frac{\xi(l_i)}{B(L)}\right) + \frac{5k^2}{6}\sum_{i=1}^{p}\frac{\xi^2(l_i)}{B^2(L)}.\end{aligned} \tag{2.11}$$

Let us define

$$\begin{aligned}\alpha_1(L) = &\frac{2}{B(L)}\sum_{i=1}^{p}\sum_{t>i}\frac{\xi(l_i)(l_i-l_t)+l_t(\xi(l_i)-\xi(l_t))}{l_i-l_t}\\ &-\frac{(p-1)}{B(L)}\sum_{i=1}^{p}\xi(l_i)\end{aligned}$$

and

$$\alpha_2(L) = 2\sum_{i=1}^{p}l_i\frac{\partial}{\partial l_i}\left(\frac{\xi(l_i)}{B(L)}\right) + \frac{5k^2}{6}\sum_{i=1}^{p}\frac{\xi^2(l_i)}{B^2(L)}.$$

We will show that $\alpha_1(L) < 0$ and $\alpha_2(L) < 0$.

Now it follows that

$$\alpha_1(L) = \frac{2}{B(L)} \sum_{i=1}^{p} \sum_{t>i} \xi(l_i) - \frac{(p-1)}{B(L)} \sum_{i=1}^{p} \xi(l_i) + \frac{2}{B(L)} \sum_{i=1}^{p} \sum_{t>i} \frac{\{\xi(l_i) - \xi(l_t)\} l_t}{l_i - l_t}$$

$$= \frac{2}{B(L)} \sum_{i=1}^{p} (p-i)\xi(l_i) - \frac{(p-1)}{B(L)} \sum_{i=1}^{p} \xi(l_i) + \frac{2}{B(L)} \sum_{i=1}^{p} \sum_{t>i} \frac{\{\xi(l_i) - \xi(l_t)\} l_t}{l_i - l_t}$$

$$= \frac{1}{B(L)} \left[(p+1) \sum_{i=1}^{p} \xi(l_i) - 2 \sum_{i=1}^{p} i\xi(l_i) \right] + \frac{2}{B(L)} \sum_{i=1}^{p} \sum_{t>i} \frac{\{\xi(l_i) - \xi(l_t)\} l_t}{l_i - l_t}.$$

Now using Lemma 2.3 and the definition of $B(L)$ it is easy that $\alpha_1(L) < 0$. Now consider $\alpha_2(L)$. It is easy to observe, using the definition of $B(L)$, that

$$\sum_{i=1}^{p} l_i \frac{\partial}{\partial l_i} \left(\frac{\xi(l_i)}{B(L)} \right) = -\frac{p\tau(u)}{b+u} - \frac{2\tau'(u)}{b+u} u + \frac{2\tau(u)}{(b+u)^2} u. \tag{2.12}$$

Thus from (2.12), it follows that

$$\alpha_2(L) = -\frac{2p\tau(u)}{b+u} - \frac{4\tau'(u)u}{b+u} + \frac{4\tau(u)u}{(b+u)^2} + \frac{5}{6}k^2 \frac{\tau^2(u)u}{(b+u)^2}$$

$$\leq -\frac{2(p-2)\tau(u)}{b+u} + k^* \frac{\tau^2(u)u}{(b+u)^2} \quad \left(\text{since } k^* = \frac{5}{6}k^2 \right)$$

$$= -\frac{\tau(u)}{b+u} \{2(p-2) - k^*\tau(u)\} < 0 \quad (\text{by } 2.8).$$

Thus it follows that $\alpha(L) < 0$, which completes the proof of the theorem. □

REMARK 2.1. Instead of shrinking or expanding the eigenvalues towards or away from the origin as in (2.7), one can replace the origin by an arbitrary fixed point $\mu = (\mu_1, \mu_2, \ldots, \mu_p)^t$, where $\mu_1 > \mu_2 > \cdots > \mu_p$. In this case one can define $\varphi_i(L)$ as

$$\varphi_i(L) = \frac{l_i}{k} - \frac{l_i \tau(u)(\log l_i - \mu_i)}{b+u}, \qquad i = 1, 2, \ldots, p, \tag{2.13}$$

where $u = \sum_{i=1}^{p}(\log l_i - \mu_i)^2$, $b > 144(p-2)^2/25k^2$, and $\tau(u)$ is a function satisfying (2.8).

REMARK 2.2. The estimator developed in (2.7) corrects $\hat{\Sigma}_0$ positively if $l_i < 1$ and negatively if $l_i > 1$, which leads to a shrinking or expanding dependent on the magnitude of the sample eigenvalues.

A recommended way of finding an improved estimator is to choose the point μ towards which the estimator shrinks or expands, adaptively. The following theorem gives an adaptive estimator of Σ.

THEOREM 2.2. *Suppose* $\hat{\Sigma}^a = R\varphi^a(L)R^t$ *where* $\varphi^a(L)$ *is defined componentwise as*

$$\begin{aligned} &\varphi_i^a(L) = \frac{l_i}{k} - \frac{l_i \tau(u)}{b+u}\left(\log l_i - p^{-1}\sum_{i=1}^{p} \log l_i\right), \qquad i = 1, 2, \ldots, p, \\ &b > \frac{144(p-3)^2}{25k^2}, \qquad u = \sum_{i=1}^{p}\left(\log l_i - p^{-1}\sum_{i=1}^{p}\log l_i\right)^2, \end{aligned} \tag{2.14}$$

and $\tau(u)$ *is an absolutely continuous function satisfying*

$$\begin{aligned} &\text{(i) } 0 < \tau(u) < (p-3)/k^*, \qquad k^* = 5k^2/6 \\ &\text{(ii) } \tau(u) \text{ monotone nondecreasing in } u, \qquad E[\tau'(u)] < \infty. \end{aligned} \tag{2.15}$$

Then $\hat{\Sigma}^a$ *will dominate* $\hat{\Sigma}_0$ *in terms of risk if* $p \geq 4$.

PROOF. The proof is omitted because of its similarity to that of Theorem 2.1. □

3. Orthogonally invariant minimax estimators. In this section we will derive estimators and show that they are minimax by showing that the risk of the estimators is smaller than the minimax risk. The following theorem gives a simple orthogonally invariant minimax estimator that is expressible in closed form for any dimension. It has been brought to our attention by the editor that this result was also obtained by Charles Stein and presented in a series of lectures given at the University of Washington, Seattle in 1982.

THEOREM 3.1. *Consider the estimator* $\hat{\Sigma}^m = R\varphi^m(L)R^t$ *where* $\varphi^m(L)$ *is given componentwise as*

$$\varphi_i^m(L) = l_i d_i, \qquad i = 1, 2, \ldots, p. \tag{3.1}$$

Then $\hat{\Sigma}^m$ *is minimax.*

PROOF. From Stein (1977a, b), it follows that the risk of $\hat{\Sigma}^m$ is given as

$$R(\hat{\Sigma}^m, \Sigma) = E\left[2\sum_{i=1}^{p}\sum_{t>i}\frac{\varphi_i^m - \varphi_t^m}{l_i - l_t} + 2\sum_{i=1}^{p}\frac{\partial\varphi_i^m}{\partial l_i} + (k-p-1)\sum_{i=1}^{p}\frac{\varphi_i^m}{l_i} - \sum_{i=1}^{p}\log\varphi_i^m + \log|\Sigma| - p\right]$$

$$= E\left[2\sum_{i=1}^{p}\sum_{t>i}\frac{l_i(d_i - d_t)}{l_i - l_t} + 2\sum_{i=1}^{p}\sum_{t>i} d_t + 2\sum_{i=1}^{p} d_i + (k-p-1)\sum_{i=1}^{p} d_i - \sum_{i=1}^{p}\log l_i - \sum_{i=1}^{p}\log d_i + \log|\Sigma| - p\right]$$

$$(3.2) \quad \le E\left[2\sum_{i=1}^{p}\sum_{t>i}(d_i - d_t) + 2\sum_{i=1}^{p}\sum_{t>i} d_t + (k-p+1)\sum_{i=1}^{p} d_i - \log\left(\frac{|S|}{|\Sigma|}\right) - \sum_{i=1}^{p}\log d_i - p\right]\left(\text{since } d_i < d_t \text{ and } \frac{l_i}{l_i - l_t} > 1\right)$$

$$= E\left[2\sum_{i=1}^{p}(p-i)d_i + (k-p+1)\sum_{i=1}^{p} d_i - \log\left(\frac{|S|}{|\Sigma|}\right) - \sum_{i=1}^{p}\log d_i - p\right]$$

$$= \sum_{i=1}^{p}(k-p+1+2p-2i)d_i - E\sum_{i=1}^{p}\log\left(\chi^2_{k+1-i}\right) - \sum_{i=1}^{p}\log d_i - p$$

$$= p - \sum_{i=1}^{p} E\left(\log\chi^2_{k+1-i}\right) - \sum_{i=1}^{p}\log d_i - p$$

$$= -\sum_{i=1}^{p}\log d_i - \sum_{i=1}^{p} E\log\left(\chi^2_{k+1-i}\right).$$

This completes the proof of the theorem. □

Here we develop a minimax estimator of the form (2.7) by improving the estimator developed in (3.1). We will show that this new estimator is better than (in terms of risk) that given in (3.1) and hence minimax.

THEOREM 3.2. *Consider the estimator* $\hat{\Sigma}^S = R\varphi^S(L)R^t$. *If* $\varphi^S(L)$ *is given componentwise as*

$$(3.3) \qquad \varphi_i^S(L) = \varphi_i^m(L) - \frac{(l_i \log l_i)\tau(u)}{b_1 + u}, \qquad i = 1, 2, \dots, p,$$

where $\varphi_i^m(L) = l_i d_i$, $i = 1, 2, \dots, p$, $u = \sum_{i=1}^{p}\log^2 l_i$, $b_1 > 144(p-2)^2/25(k+p-1)^2$, *and* $\tau(u)$ *is a function satisfying*

(3.4)
(i) $0 < \tau(u) < 12(p-2)/5(k+p-1)^2$,

(ii) $\tau(u)$ *monotone nondecreasing in u and* $E[\tau'(u)] < \infty$;

then

$$R(\hat{\Sigma}^S, \Sigma) \le R(\hat{\Sigma}^m, \Sigma) \le R(\hat{\Sigma}^M, \Sigma)$$

and hence $\hat{\Sigma}^S$ *is minimax.*

PROOF. Define $\varphi_i^S(L) = \varphi_i^m(L) + \gamma_i(L)$ and $\gamma_i(L) = l_i \eta_i(L)$, where $\eta_i(L) = [-(\log l_i)\tau(u)]/(b_1 + u)$, $i = 1, 2, \dots, p$. By using (2.9) and Lemma 2.2 with $u = \frac{1}{2}$, one finds that the unbiased estimator of the difference in risk between $\hat{\Sigma}^S$ and $\hat{\Sigma}^m$ is $\alpha^*(L)$, where

$$\begin{aligned} \alpha^*(L) \le 2 \sum_{i=1} \sum_{t>i} & \frac{l_i \eta_i(L) - l_t \eta_t(L)}{l_i - l_t} \\ & + (k - p + 1) \sum_{i=1}^{p} \eta_i(L) - \sum_{i=1}^{p} (k + p + 1 - 2i) \eta_i(L) \qquad (3.5) \\ & + 2 \sum_{i=1}^{p} l_i \frac{\partial}{\partial l_i} \eta_i(L) + \frac{5}{6} \sum_{i=1}^{p} (k + p + 1 - 2i)^2 \eta_i^2(L). \end{aligned}$$

Let us define

$$\begin{aligned} \alpha_1^*(L) = 2 \sum_{i=1}^{p} \sum_{t>i} & \frac{l_i \eta_i(L) - l_t \eta_t(L)}{l_i - l_t} \\ & + (k - p + 1) \sum_{i=1} \eta_i(L) - \sum_{i=1}^{p} (k + p + 1 - 2i) \eta_i(L) \end{aligned}$$

and

$$\alpha_2^*(L) = 2 \sum_{i=1}^{p} l_i \frac{\partial}{\partial l_i} \eta_i(L) + \frac{5}{6} \sum_{i=1}^{p} (k + p + 1 - 2i)^2 \eta_i^2(L).$$

By arguments similar to those in the proof of Theorem 2.1, it can be shown using condition (3.4) that $\alpha_2^*(L) < 0$. Furthermore,

$$\begin{aligned} \alpha_1^*(L) &= 2 \sum_{i=1}^{p} \sum_{t>i} \frac{l_i \eta_i(L) - l_t \eta_t(L)}{l_i - l_t} + 2 \sum_{i=1}^{p} (i - p) \eta_i(L) \\ &= 2 \sum_{i=1}^{p} \sum_{t>i} \eta_i(L) + 2 \sum_{i=1}^{p} \sum_{t>i} \frac{l_t[\eta_i(L) - \eta_t(L)]}{l_i - l_t} - 2 \sum_{i=1}^{p} (p - i) \eta_i(L) \\ &= 2 \sum_{i=1}^{p} (p - i) \eta_i(L) + 2 \sum_{i=1}^{p} \sum_{t>i} \frac{l_t[\eta_i(L) - \eta_t(L)]}{l_i - l_t} - 2 \sum_{i=1}^{p} (p - i) \eta_i(L) \\ &= 2 \sum_{i=1}^{p} \sum_{t>i} \frac{l_t[\eta_i(L) - \eta_t(L)]}{l_i - l_t} < 0 \end{aligned}$$

[by similar argument as given in (2.12)]. Thus $\alpha^*(L) < 0$, which completes the proof of the theorem. □

In practice we often prefer a scale invariant estimator. A recommended way of finding a scale invariant minimax estimator is to choose the point towards which the estimator shrinks or expands adaptively. The following theorem, which is analogous to Theorem 2.2, gives an adaptive minimax estimator of Σ.

THEOREM 3.3. *Consider the estimator* $\hat{\Sigma}^A = R\varphi^A(L)R$ *where* $\varphi^A(L)$ *is defined componentwise as*

$$\varphi_i^A(L) = l_i d_i - \frac{l_i \tau(u)}{b_2 + u}\left(\log l_i - p^{-1}\sum_{i=1}^{p}\log l_i\right), \qquad i = 1,2,\dots,p, \tag{3.6}$$

$$b_2 > \frac{144(p-3)^2}{25(k+p-1)^2}, \qquad u = \sum_{i=1}^{p}\left(\log l_i - p^{-1}\sum_{i=1}^{p}\log l_i\right)^2,$$

and $\tau(u)$ *is an absolutely continuous function satisfying*

(3.7)
(i) $0 < \tau(u) < (p-3)/k^*, \qquad k^* = 5k^2/6$

(ii) $\tau(u)$ *is monotone nondecreasing in* u, *and* $E[\tau'(u)] < \infty$.

Then $\hat{\Sigma}^A$ *is minimax if* $p \geq 4$.

PROOF. The proof is again omitted because of its similarity to that of Theorem 3.2. □

REMARK 3.1. The estimators given in (3.3) and (3.6) are very simple minimax estimators; however φ_i^Ss and φ_i^As are not order preserving. Following Barlow et al. (1972) one can address this problem by performing an isotonic regression technique over φ_i^Ss and φ_i^As. See Lin and Perlman (1985) for a complete description of this modification.

4. Risk simulation study. A Monte Carlo simulation study was performed to compute the risks of our minimax estimator given in (3.3) and to compare the performances with the risk of the standard estimator $\hat{\Sigma}_0$ for $p = 3$, 5, and 6 and various values of k. The covariance matrix Σ was chosen to be diagonal in such a way that they gave a wide spectrum of eigenvalues. See for details in Dey and Srinivasan (1984) for the choice of Σ. In this study, we considered the minimax estimator (3.3) with $b_1 = 5.8(p-2)^2/(k+p-1)^2$ and $\tau(u)$ a constant c with $c = 6(p-2)/5(k+p-1)^2$. Then we computed the percentage improvements in risk of our minimax estimator (3.3) over the minimax risk.

The numerical studies indicated that the percentage improvements over the risk of $\hat{\Sigma}_0$ and over the minimax risk are both significant, except for very large k. It is observed that as k increased, the percentage improvements decreased. For $\Sigma = I$, the identity matrix, the percentage improvements were over 50%. It is also observed that most of the improvement in risk (over the minimax risk) is obtained by the estimator (3.1); the additional improvement offered by (3.3) over (3.1) is relatively small. Also, both Stein's (1975) estimator and Haff's (1982) modification offer substantially greater improvement in risk than (3.1) or (3.3) when Σ is approximately scalar multiple of identity matrix.

Acknowledgment. The authors wish to thank Mr. Subhendu Lahiri for the help in numerical computation. The authors are also grateful to the editor and the referees for suggestions that improved the presentation of this paper.

REFERENCES

BARLOW, R. E. et al. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

DEY, D. K., GHOSH, M. and SRINIVASAN, C. (1984). Simultaneous estimation of parameters under entropy loss. To appear in *J. Statist. Planning & Inference*.

DEY, D. K. and SRINIVASAN, C. (1984). Estimation of covariance matrix under Stein's loss. Technical Report, Texas Tech. University.

EATON, M. L. (1970). Some problems in covariance estimation. Technical Report No. 49, Department of Statistics, Stanford University (preliminary report).

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

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

HAFF, L. R. (1979a). Estimation of the inverse convariance matrix: Random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.* **7** 1264–1276.

HAFF, L. R. (1979b). An identity for the Wishart distribution with applications. *J. Multivariate Anal.* **9** 531–542.

HAFF, L. R. (1982). Solutions of the Euler–Lagrange equations for certain multivariate normal estimation problems. Unpublished manuscript.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In: *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1**, 361–380. Univ. of California Press.

LIN, S. P. (1977). Improved procedures for estimating a correlation matrix. Unpublished Ph.D. thesis, Department of Statistics, University of Chicago.

LIN, S. P. and PERLMAN, M. D. (1985). A Monte Carlo comparison of four estimators for a covariance matrix. *Multivariate Analy.* **6**, 411–429, ed. P. R. Krishnaiah North Holland, Amsterdam.

OLKIN, I. and SELLIAH, J. B. (1977). Estimating Covariances in a Multivariate Normal Distribution. *Statistical Decision Theory and Related Topics*. **II**, 313–326, ed. S. S. Gupta & D. Moore Academic Press, New York.

SELLIAH, J. B. (1964). Estimation and testing problems in a Wishart distribution. Technical Report No. 10, Department of Statistics, Stanford University.

STEIN, C. (1975). Rietz lecture. 39th annual meeting IMS. Atlanta, Georgia.

STEIN, C. (1977a). Unpublished notes on estimating the covariance matrix.

STEIN, C. (1977b). Lectures on the theory of estimation of many parameters. (In Russian.) In: *Studies in the Statistical Theory of Estimation, Part I* (Ibragimov, I. A. and Nikulin, M. S., eds.), *Proceedings of Scientific Seminars of the Steklov Institute, Leningrad Division* **74**, 4–65.

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

WILKINSON, J. H. (1965). *The Algebraic Eigenvalue Problem*. Clarendon, Oxford.

DEPARTMENT OF MATHEMATICS
TEXAS TECH UNIVERSITY
LUBBOCK, TX 79409

DEPARTMENT OF STATISTICS
UNIVERSITY OF KENTUCKY
LEXINGTON, KENTUCKY 40506