The Annals of Statistics
1991, Vol. 19, No. 1, pp. 283–296

ESTIMATING COVARIANCE MATRICES

BY WEI-LIEM LOH

Purdue University

Let S_1 and S_2 be two independent $p \times p$ Wishart matrices with $S_1 \sim W_p(\Sigma_1, n_1)$ and $S_2 \sim W_p(\Sigma_2, n_2)$. We wish to estimate (Σ_1, Σ_2) under the loss function $L(\hat{\Sigma}_1, \hat{\Sigma}_2; \Sigma_1, \Sigma_2) = \Sigma_i\{\mathrm{tr}(\Sigma_i^{-1}\hat{\Sigma}_i) - \log|\Sigma_i^{-1}\hat{\Sigma}_i| - p\}$. Our approach is to first utilize the principle of invariance to narrow the class of estimators under consideration to the equivariant ones. The unbiased estimates of risk of these estimators are then computed and promising estimators are derived from them. A Monte Carlo study is also conducted to evaluate the risk performances of these estimators. The results of this paper extend those of Stein, Haff, Dey and Srinivasan from the one sample problem to the two sample one.

1. Introduction. A great deal of effort has been expended on constructing minimax estimators for a covariance matrix Σ of a multivariate normal distribution with the aim of getting substantial savings in risk when the eigenvalues of Σ are close together. The literature includes Stein (1975, 1977), Haff (1980, 1982, 1988) and Dey and Srinivasan (1985, 1986). A more complete list of references can be found in Loh (1988). In this paper, the two sample analogue is examined. Namely, we consider the minimax estimation of two covariance matrices (Σ_1, Σ_2) of two multivariate normal populations with the aim of getting substantial savings in risk when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are close together. For example, this would be useful in estimating (Σ_1, Σ_2) when one has prior information that the eigenvalues of Σ_i, $i = 1, 2$, are likely to be far apart and the Σ_i's are approximately proportional.

We shall use the following notation throughout. If a matrix A has entries a_{ij}, we shall indicate it by (a_{ij}). Given an $r \times s$ matrix A, its $s \times r$ transpose is denoted by A'. $|A|$ and A^{-1} denote the determinant and the inverse of the square matrix A, respectively. The trace of A is indicated by $\mathrm{tr}\, A$ and I denotes the identity matrix. If the $p \times p$ matrix A is diagonal and has entries a_{ij}, we shall write it as $A = \mathrm{diag}(a_{11}, \dots, a_{pp})$. Finally, the expected value of a random vector X is denoted by EX.

The precise formulation of the problem is as follows: Let S_1 and S_2 be two independent $p \times p$ Wishart matrices where $S_1 \sim W_p(\Sigma_1, n_1)$ and $S_2 \sim W_p(\Sigma_2, n_2)$. We wish to estimate (Σ_1, Σ_2) under the loss function:

$$(1) \qquad L(\hat{\Sigma}_1, \hat{\Sigma}_2; \Sigma_1, \Sigma_2) = \sum_{i=1}^{2} \{\mathrm{tr}(\Sigma_i^{-1}\hat{\Sigma}_i) - \log|\Sigma_i^{-1}\hat{\Sigma}_i| - p\}.$$

This loss function is convex and is the natural extension of Stein's loss in the

Received August 1988; revised August 1989.

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

Key words and phrases. Covariance matrix, Wishart distribution, Stein's loss, unbiased estimate of risk, equivariant estimation.

one sample case, where Stein's loss is given by

$$L_S(\hat{\Sigma}; \Sigma) = \operatorname{tr}(\Sigma^{-1}\hat{\Sigma}) - \log|\Sigma^{-1}\hat{\Sigma}| - p. \tag{2}$$

The loss function L_S was first considered by Stein (1956).

We shall first state the Wishart identity which was proved independently by Stein (1975) and Haff (1977). A proof of this identity can be found in Loh (1988).

A function g: $R^{p\times n} \to R$ is almost differentiable if, for every direction, the restrictions to almost all lines in that direction are absolutely continuous. If g on $R^{p\times n}$ is vector-valued instead of being real-valued, then g is almost differentiable if each of its coordinate functions are.

Let $\mathscr{S}_p$ denote the set of $p \times p$ positive definite matrices. Also for $1 \leq i \leq 2$, $1 \leq j, k \leq p$, we write $S_i = (s_{jk}^{(i)})$ and

$$\tilde{\nabla}^{(i)} = \left(\tilde{\nabla}_{jk}^{(i)}\right)_{p\times p}, \quad \text{where} \quad \tilde{\nabla}_{jk}^{(i)} = \tfrac{1}{2}(1 + \delta_{jk})\,\partial/\partial s_{jk}^{(i)},$$

where δ_{jk} denotes the Kronecker delta.

THEOREM 1 (Wishart identity). *Let $X = (X_1, \dots, X_n)$ be a $p \times n$ random matrix, with the X_k independently normally distributed p-dimensional random vectors with mean 0 and unknown covariance matrix Σ_i. We suppose $n \geq p$. Let g: $\mathscr{S}_p \to R^{p\times p}$ be such that $x \mapsto g(xx')$: $R^{p\times n} \to R^{p\times p}$ is almost differentiable. Then, with $S_i = XX'$, we have*

$$E \operatorname{tr} \Sigma_i^{-1} g(S_i) = E \operatorname{tr}\left[(n - p - 1) S_i^{-1} g(S_i) + 2\tilde{\nabla}^{(i)} g(S_i)\right],$$

provided the expectations of the two terms on the r.h.s. exist.

2. Unbiased estimate of risk. The problem which we are concerned with is invariant under the following group of transformations:

$$\Sigma_i \to A\Sigma_i A', \qquad S_i \to AS_iA' \qquad \forall A \in GL(p, R), \qquad i = 1, 2, \tag{3}$$

where GL(p, R) denotes the set of all $p \times p$ nonsingular matrices.

THEOREM 2. *Let $S_1 \sim W_p(\Sigma_1, n_1)$ and $S_2 \sim W_p(\Sigma_2, n_2)$ with S_1, S_2 independent. Then under the group of transformations given in* (3), $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ *is an equivariant estimator of (Σ_1, Σ_2) if and only if $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ can be expressed as*

$$\hat{\Sigma}_1(S_1, S_2, n_1, n_2) = B^{-1}\Psi(I - F, n_1, n_2) B'^{-1},$$

$$\hat{\Sigma}_2(S_1, S_2, n_1, n_2) = B^{-1}\Phi(F, n_2, n_1) B'^{-1},$$

where Φ, Ψ are both diagonal matrices, $B(S_1 + S_2)B' = I$, $BS_2B' = F$ and $f_1 \geq \cdots \geq f_p$ with $F = \operatorname{diag}(f_1, \dots, f_p)$.

PROOF. Suppose $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ is an equivariant estimator of (Σ_1, Σ_2). Then

$$\hat{\Sigma}_i(S_1, S_2, n_1, n_2) = A^{-1}\hat{\Sigma}_i(AS_1A', AS_2A', n_1, n_2) A'^{-1}, \tag{4}$$

for all $A \in$ GL(p, R). We observe that $\exists\ B \in$ GL(p, R) such that $B(S_1 +$

$S_2)B' = I$ and $BS_2B' = F$, where $F = \mathrm{diag}(f_1, \dots, f_p)$ with $f_1 \geq \cdots \geq f_p$. Hence it follows from (4) that

$$\hat{\Sigma}_i(S_1, S_2, n_1, n_2) = B^{-1}\hat{\Sigma}_i(I - F, F, n_1, n_2)B'^{-1}.$$

By invariance again, we have

$$\hat{\Sigma}_i(I - F, F, n_1, n_2) = D\hat{\Sigma}_i(I - F, F, n_1, n_2)D, \qquad \forall\, D = \mathrm{diag}(\pm 1).$$

This implies that $\hat{\Sigma}_i(I - F, F, n_1, n_2)$ is diagonal for $i = 1, 2$. Writing

$$\Psi(I - F, n_1, n_2) = \hat{\Sigma}_1(I - F, F, n_1, n_2),$$
$$\Phi(F, n_2, n_1) = \hat{\Sigma}_2(I - F, F, n_1, n_2),$$

proves the necessity part. For the sufficiency part of the result, the proof is straightforward and is omitted. □

Next we state a rather technical lemma. Its proof, which uses the calculus on eigenstructure techniques of Stein (1975) and Haff (1988), is tedious and we refer the reader to Loh (1988) for details.

LEMMA 1. *Let $B, \Phi = \mathrm{diag}(\phi_1, \dots, \phi_p)$ and $\Psi = \mathrm{diag}(\psi_1, \dots, \psi_p)$ be defined as in Theorem 2. Then*

$$\mathrm{tr}\,\tilde{\nabla}^{(1)}(B^{-1}\Psi B'^{-1}) = \sum_i \left[\psi_i + f_i \frac{\partial \psi_i}{\partial(1 - f_i)} + \psi_i \sum_{j \neq i} \frac{f_j}{f_j - f_i}\right],$$
$$\mathrm{tr}\,\tilde{\nabla}^{(2)}(B^{-1}\Phi B'^{-1}) = \sum_i \left[\phi_i + (1 - f_i)\frac{\partial \phi_i}{\partial f_i} + \phi_i \sum_{j \neq i} \frac{1 - f_j}{f_j - f_i}\right].$$

The next two propositions follow from the Wishart identity and Lemma 1.

PROPOSITION 1. *Let $\hat{\Sigma}_1$ be an estimator for Σ_1 where*

$$\hat{\Sigma}_1(S_1, S_2, n_1, n_2) = B^{-1}\Psi(I - F, n_1, n_2)B'^{-1},$$

$\Psi = \mathrm{diag}(\psi_1, \dots, \psi_p)$, $B(S_1 + S_2)B' = I$ and $BS_2B' = F = \mathrm{diag}(f_1, \dots, f_p)$ with $f_1 \geq \cdots \geq f_p$. Suppose Ψ satisfies the conditions of the Wishart identity in the sense that $E\,\mathrm{tr}(\Sigma_1^{-1}\hat{\Sigma}_1) = E\,\mathrm{tr}[2\tilde{\nabla}^{(1)}(\hat{\Sigma}_1) + (n_1 - p - 1)S_1^{-1}\hat{\Sigma}_1]$. Then under Stein's loss, the risk of $\hat{\Sigma}_1$ is given by

$$R_S(\hat{\Sigma}_1; \Sigma_1) = E\left\{\sum_i \left[\frac{n_1 - p - 1}{1 - f_i}\psi_i - 2\psi_i \sum_{j \neq i} \frac{f_j}{f_i - f_j} + 2\psi_i \right.\right.$$
$$\left.\left. + 2 f_i \frac{\partial \psi_i}{\partial(1 - f_i)} - \log\frac{\psi_i}{1 - f_i} - \log \chi^2_{n_1 - i + 1} - 1\right]\right\}.$$

PROPOSITION 2. *Let $\hat{\Sigma}_2$ be an estimator for Σ_2 where*

$$\hat{\Sigma}_2(S_1, S_2, n_1, n_2) = B^{-1}\Phi(F, n_2, n_1)B'^{-1},$$

$\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_p)$, $B(S_1 + S_2)B' = I$ and $BS_2B' = F = \mathrm{diag}(f_1, \ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$. Suppose Φ satisfies the conditions of the Wishart identity in the sense that $E\,\mathrm{tr}(\Sigma_2^{-1}\hat{\Sigma}_2) = E\,\mathrm{tr}[2\tilde{\nabla}^{(2)}(\hat{\Sigma}_2) + (n_1 - p - 1)S_2^{-1}\hat{\Sigma}_2]$. Then under Stein's loss, the risk of $\hat{\Sigma}_2$ is given by

$$R_S(\hat{\Sigma}_2; \Sigma_2) = E\left\{\sum_i \left[\frac{n_2 - p - 1}{f_i}\phi_i + 2\phi_i \sum_{j \neq i} \frac{1 - f_j}{f_i - f_j} + 2\phi_i + 2(1 - f_i)\frac{\partial \phi_i}{\partial f_i} - \log \frac{\phi_i}{f_i} - \log \chi^2_{n_2 - i + 1} - 1\right]\right\}.$$

An immediate consequence of Propositions 1 and 2 is:

THEOREM 3. *Let $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ be an estimator for (Σ_1, Σ_2) where*

$$\hat{\Sigma}_1(S_1, S_2, n_1, n_2) = B^{-1}\Psi(I - F, n_1, n_2)B'^{-1},$$
$$\hat{\Sigma}_2(S_1, S_2, n_1, n_2) = B^{-1}\Phi(F, n_2, n_1)B'^{-1}.$$

Then under the conditions of Propositions 1 and 2, the risk of $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ with respect to the loss function L is given by

$$\begin{aligned} R(\hat{\Sigma}_1, \hat{\Sigma}_2; \Sigma_1, \Sigma_2) &= E\left\{\sum_i \left[\frac{n_1 - p - 1}{1 - f_i}\psi_i - 2\psi_i \sum_{j \neq i} \frac{f_j}{f_i - f_j} + 2\psi_i \right.\right. \\ &\quad + 2f_i \frac{\partial \psi_i}{\partial(1 - f_i)} - \log \frac{\psi_i}{1 - f_i} + \frac{n_2 - p - 1}{f_i}\phi_i \\ &\quad + 2\phi_i \sum_{j \neq i} \frac{1 - f_j}{f_i - f_j} + 2\phi_i + 2(1 - f_i)\frac{\partial \phi_i}{\partial f_i} \\ &\quad \left.\left. - \log \frac{\phi_i}{f_i} - \log \chi^2_{n_1 - i + 1} - \log \chi^2_{n_2 - i + 1} - 2\right]\right\}. \end{aligned}$$

3. Usual estimators and minimax risk. The usual estimators $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ of (Σ_1, Σ_2) are of the form (c_1S_1, c_2S_2) where c_1, c_2 are constants. The best usual estimator is that usual estimator which minimizes the risk among the usual estimators. It is easily shown that with respect to the loss function L, the best usual estimator $(\hat{\Sigma}_1^{\mathrm{BU}}, \hat{\Sigma}_2^{\mathrm{BU}})$ of (Σ_1, Σ_2) is $(S_1/n_1, S_2/n_2)$ and that

$$R(\hat{\Sigma}_1^{\mathrm{BU}}, \hat{\Sigma}_2^{\mathrm{BU}}; \Sigma_1, \Sigma_2) = E\left[p \log n_1 + p \log n_2 - \sum_i \left(\log \chi^2_{n_1 - i + 1} + \log \chi^2_{n_2 - i + 1}\right)\right]. \tag{5}$$

Next we state a one sample minimax result of Stein (1956).

THEOREM 4. *Let $S \sim W_p(\Sigma, n)$. With respect to Stein's loss, the best estimator equivariant with respect to linear transformations $\Sigma \to U\Sigma U'$, $S \to USU'$, where U is nonsingular lower triangular, is $\hat{\Sigma}^{\mathrm{MM}}(S) = TDT'$, where the jth diagonal element of the diagonal matrix D is $1/(n + p - 2j + 1)$, $j = 1, \dots, p$, and $S = TT'$, with T lower triangular. This estimator is minimax with risk*

$$R_S(\hat{\Sigma}^{\mathrm{MM}}; \Sigma) = E\left[\sum_j \log(n + p - 2j + 1) - \sum_j \log \chi^2_{n-j+1}\right].$$

We shall now give a two sample analogue of Theorem 4. To do so, we shall consider the class of equivariant estimators of (Σ_1, Σ_2) under the group of transformations

$$\Sigma_i \to U_i \Sigma_i U_i', \qquad S_i \to U_i S_i U_i', \qquad i = 1, 2, \tag{6}$$

where U_i is a nonsingular lower triangular matrix.

THEOREM 5. *Let $S_1 \sim W_p(\Sigma_1, n_1)$ and $S_2 \sim W_p(\Sigma_2, n_2)$ with S_1 and S_2 independent. With respect to the loss function L, the best estimator equivariant under the group of transformations (6) is $(\hat{\Sigma}_1^{\mathrm{MM}}, \hat{\Sigma}_2^{\mathrm{MM}}) = (T_1 D_1 T_1', T_2 D_2 T_2')$, where, for $i = 1, 2$, the jth diagonal element of the diagonal matrix D_i is $1/(n_i + p - 2j + 1)$ and $S_i = T_i T_i'$, with T_i lower triangular. This estimator is minimax and has constant risk given by*

$$R\left(\hat{\Sigma}_1^{\mathrm{MM}}, \hat{\Sigma}_2^{\mathrm{MM}}; \Sigma_1, \Sigma_2\right) = E\left\{\sum_{i=1}^{2}\left[\sum_j \log(n_i + p - 2j + 1) - \sum_j \log \chi^2_{n_i-j+1}\right]\right\}.$$

The proof of Theorem 5 is similar to its one sample counterpart and we refer the reader to Loh (1988) for details. We observe from (5) and Theorem 5 that $(\hat{\Sigma}_1^{\mathrm{BU}}, \hat{\Sigma}_2^{\mathrm{BU}})$ is not minimax and that $(\hat{\Sigma}_1^{\mathrm{MM}}, \hat{\Sigma}_2^{\mathrm{MM}})$ dominates $(\hat{\Sigma}_1^{\mathrm{BU}}, \hat{\Sigma}_2^{\mathrm{BU}})$. This implies that in evaluating the risk performance of an alternative estimator for (Σ_1, Σ_2), the estimator to compare with is $(\hat{\Sigma}_1^{\mathrm{MM}}, \hat{\Sigma}_2^{\mathrm{MM}})$, not $(\hat{\Sigma}_1^{\mathrm{BU}}, \hat{\Sigma}_2^{\mathrm{BU}})$.

4. Alternative estimators. It is well known that the eigenvalues of $S_2(S_1 + S_2)^{-1}$ are more spread out than the eigenvalues of its expectation. By correcting for this eigenvalue distortion, we construct alternative estimators which compare favorably with the constant risk minimax estimator $(\hat{\Sigma}_1^{\mathrm{MM}}, \hat{\Sigma}_2^{\mathrm{MM}})$. Furthermore, these estimators give substantial savings in risk when the eigenvalues of $\Sigma_2 \Sigma_1^{-1}$ are close together.

4.1. *Adjusted usual estimator.* The best usual estimator $(\hat{\Sigma}_1^{\mathrm{BU}}, \hat{\Sigma}_2^{\mathrm{BU}})$ can be written as

$$\left(\hat{\Sigma}_1^{\mathrm{BU}}, \hat{\Sigma}_2^{\mathrm{BU}}\right) = \left(B^{-1}\Psi^{\mathrm{BU}}B'^{-1}, B^{-1}\Phi^{\mathrm{BU}}B'^{-1}\right),$$

where the jth diagonal element of the diagonal matrix Ψ^{BU} and Φ^{BU} is $(1 - f_j)/n_1$ and f_j/n_2, respectively. A natural way to improve on this estimator would be to consider estimators of the form

$$(\hat{\Sigma}_1, \hat{\Sigma}_2) = (B^{-1}\Psi B'^{-1}, B^{-1}\Phi B'^{-1}),$$

where for some constants c_j, d_j, $j = 1, \dots, p$, the jth diagonal element of the diagonal matrix Ψ and Φ is $\psi_j = (1 - f_j)/c_j$ and $\phi_j = f_j/d_j$, respectively. We define the adjusted usual estimator to be

$$(\hat{\Sigma}_1^{\mathrm{AU}}, \hat{\Sigma}_2^{\mathrm{AU}}) = (B^{-1}\Psi^{\mathrm{AU}}B'^{-1}, B^{-1}\Phi^{\mathrm{AU}}B'^{-1}),$$

where, for $j = 1, \dots, p$, the jth diagonal element of the diagonal matrix Ψ^{AU} and Φ^{AU} is $\psi_j^{\mathrm{AU}} = (1 - f_j)/(n_1 - p - 1 + 2j)$ and $\phi_j^{\mathrm{AU}} = f_j/(n_2 + p + 1 - 2j)$, respectively.

PROPOSITION 3. *Under Stein's loss, $\hat{\Sigma}_i^{\mathrm{AU}}$ is a minimax estimator of Σ_i, $i = 1, 2$.*

PROOF. This follows easily from the comparison of the unbiased estimate of the risk of $\hat{\Sigma}_i^{\mathrm{AU}}$ (which can be obtained from Propositions 1 and 2) and the minimax risk. □

THEOREM 6. *With respect to the loss function L, $(\hat{\Sigma}_1^{\mathrm{AU}}, \hat{\Sigma}_2^{\mathrm{AU}})$ dominates $(\hat{\Sigma}_1^{\mathrm{MM}}, \hat{\Sigma}_2^{\mathrm{MM}})$. Hence $(\hat{\Sigma}_1^{\mathrm{AU}}, \hat{\Sigma}_2^{\mathrm{AU}})$ is a minimax estimator of (Σ_1, Σ_2).*

PROOF. This theorem follows directly from Theorem 5 and Proposition 3 since the loss function under consideration is the sum of the respective loss functions of these two problems. □

We remark that the one sample analogue of the adjusted usual estimator was obtained independently by Stein (1975) and Dey and Srinivasan (1985).

4.2. *Dey–Srinivasan-type estimators.* In the estimation of a covariance matrix, Dey and Srinivasan (1985) constructed a class of minimax estimators for Σ. In this subsection, we shall derive an analogous class of minimax estimators for (Σ_1, Σ_2) with the aim of achieving substantial savings in risk when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are close together. First we need some additional notation. We let

$$\hat{\Sigma}_1^{\mathrm{DS}} = B^{-1}\Psi^{\mathrm{DS}}B'^{-1}, \qquad \hat{\Sigma}_2^{\mathrm{DS}} = B^{-1}\Phi^{\mathrm{DS}}B'^{-1},$$

where $\Psi^{DS} = \mathrm{diag}(\psi_1^{DS}, \ldots, \psi_p^{DS})$ and $\Phi^{DS} = \mathrm{diag}(\phi_1^{DS}, \ldots, \phi_p^{DS})$ with

$$\psi_i^{DS} = \frac{1 - f_i}{n_1 - p - 1 + 2i} + (1 - f_i)\beta_i,$$

$$\phi_i^{DS} = \frac{f_i}{n_2 + p + 1 - 2i} + f_i\gamma_i,$$

$$u = \sum_j \log^2\left(\frac{f_j}{1 - f_j}\right),$$

$$\beta_i = \left[a(u)\log\left(\frac{f_i}{1 - f_i}\right)\right] \Big/ (c + u),$$

$$\gamma_i = \left[b(u)\log\left(\frac{1 - f_i}{f_i}\right)\right] \Big/ (d + u),$$

a, b: $R^+ \to R$ being suitable functions and c, d being suitable constants.

Proposition 4. *In the estimation of Σ_1 under Stein's loss, $\hat{\Sigma}_1^{DS}$ dominates $\hat{\Sigma}_1^{AU}$ whenever*

(i) $a(u) \geq 0$ *and* $a'(u) \geq 0$ *for all* $u \geq 0$,
(ii) $\sup_{u \geq 0} a(u)(n_1 + p - 1)/(2\sqrt{c}) = x < 1$,
(iii) $\sup_{u \geq 0} a(u)(3 - x)/[6(1 - x)] \leq 2(p - 2)/(n_1 + p - 1)^2$.

Hence $\hat{\Sigma}_1^{DS}$ is minimax.

Proof. First we note that for $i = 1, \ldots, p$ and $u \geq 0$,

$$|(n_1 - p - 1 + 2i)\beta_i| \leq a(u)(n_1 + p - 1)/(2\sqrt{c})$$

and

$$\begin{aligned} &-\log[1 + (n_1 - p - 1 + 2i)\beta_i] \\ &\quad \leq (n_1 - p - 1 + 2i)^2\beta_i^2(3 - x)/[6(1 - x)] - (n_1 - p - 1 + 2i)\beta_i. \end{aligned} \tag{7}$$

Hence we observe from Proposition 1 and (7) that the difference in risk between $\hat{\Sigma}_1^{DS}$ and $\hat{\Sigma}_1^{AU}$ satisfies

$$\begin{aligned} &R_S(\hat{\Sigma}_1^{DS}; \Sigma_1) - R_S(\hat{\Sigma}_1^{AU}; \Sigma_1) \\ &\quad \leq E\sum_i \Bigg[(2 - 2i)\beta_i + 2\sum_{j<i} \frac{f_j(1 - f_i)\beta_i - f_i(1 - f_j)\beta_j}{f_j - f_i} \\ &\qquad -2f_i(1 - f_i)\frac{\partial\beta_i}{\partial f_i} + \frac{3 - x}{6(1 - x)}(n_1 - p - 1 + 2i)^2\beta_i^2\Bigg]. \end{aligned} \tag{8}$$

Next we observe that

$$(9) \qquad \sum_i \left[(2-2i)\beta_i + 2\sum_{j<i} \frac{f_j(1-f_i)\beta_i - f_i(1-f_j)\beta_j}{f_j - f_i} \right] \le 0,$$

$$(10) \qquad \sum_i \left[-2f_i(1-f_i)\frac{\partial \beta_i}{\partial f_i} + \frac{3-x}{6(1-x)}(n_1 - p - 1 + 2i)^2 \beta_i^2 \right] \le 0.$$

We now conclude from (8), (9) and (10) that $\hat{\Sigma}_1^{\mathrm{DS}}$ dominates $\hat{\Sigma}_1^{\mathrm{AU}}$. Minimaxity follows from Theorem 6. □

The proof of the next proposition is similar to Proposition 4 and is omitted.

PROPOSITION 5. *In the estimation of* Σ_2 *under Stein's loss,* $\hat{\Sigma}_2^{\mathrm{DS}}$ *dominates* $\hat{\Sigma}_2^{\mathrm{AU}}$ *whenever*

(i) $b(u) \ge 0$ *and* $b'(u) \ge 0$ *for all* $u \ge 0$,
(ii) $\sup_{u \ge 0} b(u)(n_2 + p - 1)/(2\sqrt{d}) = y < 1$,
(iii) $\sup_{u \ge 0} b(u)(3-y)/[6(1-y)] \le 2(p-2)/(n_2+p-1)^2$.

Hence $\hat{\Sigma}_2^{\mathrm{DS}}$ *is minimax.*

An immediate consequence of the above two propositions is

THEOREM 7. *In the estimation of* (Σ_1, Σ_2) *under the loss function* L, $(\hat{\Sigma}_1^{\mathrm{DS}}, \hat{\Sigma}_2^{\mathrm{DS}})$ *dominates* $(\hat{\Sigma}_1^{\mathrm{AU}}, \hat{\Sigma}_2^{\mathrm{AU}})$ *whenever*

(i) $a(u) \ge 0$ *and* $a'(u) \ge 0$ *for all* $u \ge 0$,
(ii) $b(u) \ge 0$ *and* $b'(u) \ge 0$ *for all* $u \ge 0$,
(iii) $\sup_{u \ge 0} a(u)(n_1 + p - 1)/(2\sqrt{c}) = x < 1$,
(iv) $\sup_{u \ge 0} b(u)(n_2 + p - 1)/(2\sqrt{d}) = y < 1$,
(v) $\sup_{u \ge 0} a(u)(3-x)/[6(1-x)] \le 2(p-2)/(n_1+p-1)^2$,
(vi) $\sup_{u \ge 0} b(u)(3-y)/[6(1-y)] \le 2(p-2)/(n_2+p-1)^2$.

Hence $(\hat{\Sigma}_1^{\mathrm{DS}}, \hat{\Sigma}_2^{\mathrm{DS}})$ *is minimax.*

Analogous to Dey and Srinivasan (1985), one can construct adapted versions of these minimax estimators. For details, we refer the reader to Loh (1988).

4.3. *Stein-type estimator.* By an approximate minimization of the unbiased estimate of the risk of an almost arbitrary orthogonally invariant estimator of a covariance matrix, Stein (1975) constructed an estimator whose risk compares very favorably with the minimax risk. In particular, substantial savings in risk is obtained when the eigenvalues of the population covariance matrix are close together.

In this subsection, this technique is applied to construct an alternative equivariant estimator $(\hat{\Sigma}_1^{\mathrm{ST}}, \hat{\Sigma}_2^{\mathrm{ST}})$ for (Σ_1, Σ_2). Let $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ be an estimator for

(Σ_1, Σ_2) where

$$\hat{\Sigma}_1(S_1, S_2, n_1, n_2) = B^{-1}\Psi(I - F, n_1, n_2)B'^{-1},$$
$$\hat{\Sigma}_2(S_1, S_2, n_1, n_2) = B^{-1}\Phi(F, n_2, n_1)B'^{-1},$$

$\Phi = \operatorname{diag}(\phi_1, \dots, \phi_p)$, $\Psi = \operatorname{diag}(\psi_1, \dots, \psi_p)$, $B(S_1 + S_2)B' = I$ and $BS_2B' = F = \operatorname{diag}(f_1, \dots, f_p)$ with $f_1 \geq \cdots \geq f_p$. Under loss function L, we observe from Theorem 3 that

$$\begin{aligned} R(\hat{\Sigma}_1, \hat{\Sigma}_2; \Sigma_1, \Sigma_2) = E\Bigg\{ \sum_i \Bigg[& \frac{n_1 - p + 1}{1 - f_i}\psi_i - 2\psi_i \sum_{j \neq i} \frac{f_j}{f_i - f_j} \\ & + 2f_i(1 - f_i)\frac{\partial}{\partial(1 - f_i)}\left(\frac{\psi_i}{1 - f_i}\right) - \log\frac{\psi_i}{1 - f_i} \\ & - \log \chi^2_{n_1 - i + 1} + \frac{n_2 - p + 1}{f_i}\phi_i + 2\phi_i \sum_{j \neq i} \frac{1 - f_j}{f_i - f_j} \\ & + 2f_i(1 - f_i)\frac{\partial}{\partial f_i}\left(\frac{\phi_i}{f_i}\right) - \log\frac{\phi_i}{f_i} - \log \chi^2_{n_2 - i + 1} - 2 \Bigg] \Bigg\}. \end{aligned}$$

By ignoring the derivative terms in the unbiased estimate of the risk, we get

$$\begin{aligned} \hat{R} = \sum_i \Bigg[& \frac{n_1 - p + 1}{1 - f_i}\psi_i - 2\psi_i \sum_{j \neq i} \frac{f_j}{f_i - f_j} - \log\frac{\psi_i}{1 - f_i} \\ & - E \log \chi^2_{n_1 - i + 1} + \frac{n_2 - p + 1}{f_i}\phi_i + 2\phi_i \sum_{j \neq i} \frac{1 - f_j}{f_i - f_j} \\ & - \log\frac{\phi_i}{f_i} - E \log \chi^2_{n_2 - i + 1} - 2 \Bigg]. \end{aligned}$$

Now we minimize $\hat{R}$ with respect to ψ_i and ϕ_i, $i = 1, \dots, p$. This gives

$$\partial\hat{R}/\partial\psi_i = 0, \qquad \partial\hat{R}/\partial\phi_i = 0, \qquad \forall\, i.$$

On simplification, we have

$$\begin{aligned} \psi_i &= (1 - f_i) \Bigg/ \left[n_1 - p + 1 - 2\sum_{j \neq i} \frac{f_j(1 - f_i)}{f_i - f_j} \right], \\ \phi_i &= f_i \Bigg/ \left[n_2 - p - 1 + 2\sum_{j \neq i} \frac{f_i(1 - f_j)}{f_i - f_j} \right], \qquad i = 1, \dots, p. \end{aligned} \tag{11}$$

We observe that the ψ_i's and ϕ_i's should follow a natural ordering:

$$0 \leq \psi_1 \leq \cdots \leq \psi_p, \qquad \phi_1 \geq \cdots \geq \phi_p \geq 0.$$

However with the ψ_i's and ϕ_i's defined by (11), this ordering may be altered. By applying Stein's isotonic regression [Stein (1975)] to these ψ_i's and ϕ_i's, we arrive at a new set of ψ_i's and ϕ_i's, denoted by ψ_i^{ST} and ϕ_i^{ST}, $i = 1, \dots, p$, which satisfy

$$0 \le \psi_1^{\mathrm{ST}} \le \cdots \le \psi_p^{\mathrm{ST}}, \qquad \phi_1^{\mathrm{ST}} \ge \cdots \ge \phi_p^{\mathrm{ST}} \ge 0.$$

For a detailed description of Stein's isotonic regression, see for example Lin and Perlman (1985). We now define

$$\hat{\Sigma}_1^{\mathrm{ST}} = B^{-1}\Psi^{\mathrm{ST}}(I - F, n_1, n_2)B'^{-1},$$
$$\hat{\Sigma}_2^{\mathrm{ST}} = B^{-1}\Phi^{\mathrm{ST}}(F, n_2, n_1)B'^{-1},$$

where $\Phi^{\mathrm{ST}} = \mathrm{diag}(\phi_1^{\mathrm{ST}}, \dots, \phi_p^{\mathrm{ST}})$ and $\Psi^{\mathrm{ST}} = \mathrm{diag}(\psi_1^{\mathrm{ST}}, \dots, \psi_p^{\mathrm{ST}})$. This concludes our construction of the Stein-type estimator.

4.4. *Haff-type estimator.* Haff (1988) constructed an estimator for a covariance matrix which has a similar functional form to that of the Stein (1975) estimator. Here we apply Haff's method to obtain an alternative estimator, denoted by $(\hat{\Sigma}_1^{\mathrm{HF}}, \hat{\Sigma}_2^{\mathrm{HF}})$, for (Σ_1, Σ_2). We note that an equivariant estimator $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ for (Σ_1, Σ_2) must be of the form

$$\hat{\Sigma}_1(S_1, S_2, n_1, n_2) = B^{-1}\Psi(I - F, n_1, n_2)B'^{-1},$$
$$\hat{\Sigma}_2(S_1, S_2, n_1, n_2) = B^{-1}\Phi(F, n_2, n_1)B'^{-1},$$

where $\Phi = \mathrm{diag}(\phi_1, \dots, \phi_p)$, $\Psi = \mathrm{diag}(\psi_1, \dots, \psi_p)$, $B(S_1 + S_2)B' = I$ and $BS_2B' = F = \mathrm{diag}(f_1, \dots, f_p)$ with $f_1 \ge \cdots \ge f_p$. Since the ψ_i's and ϕ_i's follow the ordering $0 \le \psi_1 \le \cdots \le \psi_p$ and $\phi_1 \ge \cdots \ge \phi_p \ge 0$, we write for each i,

$$\psi_i(F) = \sum_{k \ge i} \varepsilon_k^2(F), \qquad \phi_i(F) = \sum_{k \ge i} \epsilon_k^2(F).$$

Hence from Theorem 3, the unbiased estimate of the risk of an almost arbitrary equivariant estimator for (Σ_1, Σ_2) can be expressed as

$$\begin{aligned}\hat{R} = \sum_i \Bigg[& \frac{n_1 - p - 1}{1 - f_i} \sum_{k \ge i} \varepsilon_k^2 - 2\sum_{k \ge i} \varepsilon_k^2 \sum_{j \ne i} \frac{f_j}{f_i - f_j} + 2\sum_{k \ge i} \varepsilon_k^2 \\ & + 2 f_i \sum_{k \ge i} \frac{\partial \varepsilon_k^2}{\partial(1 - f_i)} - \log\left(\sum_{k \ge i} \frac{\varepsilon_k^2}{1 - f_i} \right) \\ & + \frac{n_2 - p - 1}{f_i} \sum_{k \ge i} \epsilon_k^2 + 2\sum_{k \ge i} \epsilon_k^2 \sum_{j \ne i} \frac{1 - f_j}{f_i - f_j} + 2\sum_{k \ge i} \epsilon_k^2 \\ & + 2(1 - f_i) \sum_{k \ge i} \frac{\partial \epsilon_k^2}{\partial f_i} - \log\left(\sum_{k \ge i} \frac{\epsilon_k^2}{f_i} \right) \\ & - E \log \chi_{n_1 - i + 1}^2 - E \log \chi_{n_2 - i + 1}^2 - 2 \Bigg].\end{aligned}$$

Next we put a prior distribution on the parameter space $\{(\Sigma_1, \Sigma_2)$: Σ_1, Σ_2 being positive definite matrices} and let $m(F)$ denote the marginal density of F. The average risk of this estimator is

$$\int \hat{G}\left(f_1,\ldots,f_p;\psi_1,\ldots,\psi_p;\phi_1,\ldots,\phi_p;\frac{\partial\psi_1}{\partial f_1},\ldots,\frac{\partial\psi_p}{\partial f_p};\frac{\partial\phi_1}{\partial f_1},\ldots,\frac{\partial\phi_p}{\partial f_p}\right) dF,$$

where $\hat{G} = m\hat{R}$. The solution of the Euler–Lagrange equations minimizes the average risk. These equations are

$$\hat{G}_{\psi_i} = \sum_j \frac{\partial}{\partial f_j}\hat{G}_{\partial\psi_i/\partial f_j}, \qquad \hat{G}_{\phi_i} = \sum_j \frac{\partial}{\partial f_j}\hat{G}_{\partial\phi_i/\partial f_j}, \qquad \forall\, i = 1,\ldots,p,$$

where $\hat{G}_{\psi_i} = \partial\hat{G}/\partial\psi_i$, etc. Evaluating the above set of equations for each k, $1 \le k \le p$, we have

$$(12)\qquad \begin{aligned} &\varepsilon_k \sum_{i\le k}\left[\frac{n_1-p-1}{1-f_i} - 2\sum_{j\ne i}\frac{f_j}{f_i-f_j} + 4 - \psi_i^{-1} + 2f_i\frac{\partial \log m}{\partial f_i}\right] = 0,\\ &\epsilon_k \sum_{i\le k}\left[\frac{n_2-p-1}{f_i} + 2\sum_{j\ne i}\frac{1-f_j}{f_i-f_j} + 4 - \phi_i^{-1} - 2(1-f_i)\frac{\partial \log m}{\partial f_i}\right] = 0. \end{aligned}$$

Next we set $m(F) = \prod_i 1/[f_i(1-f_i)]$. This is motivated by the observation that in order for the estimator $(\hat{\Sigma}_1^{\mathrm{HF}}, \hat{\Sigma}_2^{\mathrm{HF}})$ to compete favorably with $(\hat{\Sigma}_1^{\mathrm{MM}}, \hat{\Sigma}_2^{\mathrm{MM}})$ and $(\hat{\Sigma}_1^{\mathrm{AU}}, \hat{\Sigma}_2^{\mathrm{AU}})$, the form of $(\hat{\Sigma}_1^{\mathrm{HF}}, \hat{\Sigma}_2^{\mathrm{HF}})$ should approach that of $(\hat{\Sigma}_1^{\mathrm{AU}}, \hat{\Sigma}_2^{\mathrm{AU}})$ when the eigenvalues of $\Sigma_2(\Sigma_1+\Sigma_2)^{-1}$ are far apart. Thus (12) simplifies to

$$\begin{aligned} &\varepsilon_k \sum_{i\le k}\left[\frac{n_1-p+1}{1-f_i} - 2\sum_{j\ne i}\frac{f_j}{f_i-f_j} - \psi_i^{-1}\right] = 0,\\ &\epsilon_k \sum_{i\le k}\left[\frac{n_2-p+1}{f_i} + 2\sum_{j\ne i}\frac{1-f_j}{f_i-f_j} - \phi_i^{-1}\right] = 0. \end{aligned}$$

These equations can be solved by using an algorithm due to Haff. For a detailed description of Haff's algorithm, see Haff (1988). We denote the solution of the above equations by $\varepsilon_i^{\mathrm{HF}}$ and ϵ_i^{HF}, $1 \le i \le p$, and write

$$\psi_i^{\mathrm{HF}}(F) = \sum_{k\ge i}\varepsilon_k^{\mathrm{HF}^2}(F), \qquad \phi_i^{\mathrm{HF}}(F) = \sum_{k\ge i}\epsilon_k^{\mathrm{HF}^2}(F).$$

We now define

$$\begin{aligned} \hat{\Sigma}_1^{\mathrm{HF}} &= B^{-1}\Psi^{\mathrm{HF}}(I-F,n_1,n_2)B'^{-1},\\ \hat{\Sigma}_2^{\mathrm{HF}} &= B^{-1}\Phi^{\mathrm{HF}}(F,n_2,n_1)B'^{-1}, \end{aligned}$$

where $\Phi^{\mathrm{HF}} = \mathrm{diag}(\phi_1^{\mathrm{HF}},\ldots,\phi_p^{\mathrm{HF}})$ and $\Psi^{\mathrm{HF}} = \mathrm{diag}(\psi_1^{\mathrm{HF}},\ldots,\psi_p^{\mathrm{HF}})$. This completes our construction of the Haff-type estimator for (Σ_1, Σ_2).

5. Monte Carlo study. From the rather complicated nature of the Stein-type and Haff-type estimators, it appears that an analytical treatment of the risk performances of these estimators is not possible at this point. Using Monte Carlo simulations, we shall study the risk performances of the alternative estimators for (Σ_1, Σ_2) that we have developed in previous sections. For the simulations, we take $p = 10$, $n_1 = 12, 25$ and $n_2 = 12, 25$. Independent standard normal variates are generated by the IMSL subroutine DRNNOA and the eigenvalue decomposition uses the IMSL subroutine DEVCSF. The average loss and its estimated standard deviation of each estimator for (Σ_1, Σ_2) are computed over 500 independent replications. For brevity, we write $(\hat{\Sigma}_1^{BU}, \hat{\Sigma}_2^{BU}) = \mathrm{BU}$, $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU}) = \mathrm{AU}$, etc. As it is, the estimator $(\hat{\Sigma}_1^{DS}, \hat{\Sigma}_2^{DS})$ is not well defined. In this study we take

$$a = 6(p-2)/\left[5(n_1+p-1)^2\right],$$

$$b = 6(p-2)/\left[5(n_2+p-1)^2\right],$$

$$c = 5.8(p-2)^2/(n_1+p-1)^2,$$

$$d = 5.8(p-2)^2/(n_2+p-1)^2.$$

These values are chosen with the aim of doing well when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are close together. Table 1 gives some of these simulations. We also wish to remark that in our simulations, for a fixed set of eigenvalues of $\Sigma_2\Sigma_1^{-1}$, the estimators are computed from the same set of 500 independently generated samples. This suggests that there is a high correlation among the average losses of these estimators. Since we are more interested in the relative risk ordering of these estimators, we conclude that the estimated standard deviation (as given in Table 1) is probably a conservative indicator of the variability of the relative magnitude of the average losses. For more details on the simulations, we refer the reader to Loh (1988). The results of this numerical study indicate that:

1. For the estimation of (Σ_1, Σ_2), the risk of the alternative estimators compare very favorably with the minimax risk. Maximum savings in risk are achieved when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are all equal and the savings decrease as the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ get more and more dispersed.
2. Among the estimators, $(\hat{\Sigma}_1^{ST}, \hat{\Sigma}_2^{ST})$ and $(\hat{\Sigma}_1^{HF}, \hat{\Sigma}_2^{HF})$ perform best when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are close together with risk reduction of about 30 to 40% with respect to the minimax risk. Furthermore it is worth noting that in no instance in this simulation did the average losses of $(\hat{\Sigma}_1^{ST}, \hat{\Sigma}_2^{ST})$ and $(\hat{\Sigma}_1^{HF}, \hat{\Sigma}_2^{HF})$ exceed that of $(\hat{\Sigma}_1^{MM}, \hat{\Sigma}_2^{MM})$. Also it appears that $(\hat{\Sigma}_1^{HF}, \hat{\Sigma}_2^{HF})$ has slightly smaller risk than $(\hat{\Sigma}_1^{ST}, \hat{\Sigma}_2^{ST})$.
3. Although it has been proved that $(\hat{\Sigma}_1^{DS}, \hat{\Sigma}_2^{DS})$ dominates $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$, this Monte Carlo study reveals that the difference in risk between these two estimators is rather small at best. However as pointed out by a referee, the significance of the Dey–Srinivasan-type estimators is not in achieving sub-

TABLE 1
$n_1 = 12 \quad n_2 = 12$
Average losses of estimators for the estimation of (Σ_1, Σ_2)
(estimated standard errors are in parentheses)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	BU	MM	AU	DS	ST	HF
(1, 1, 1, 1, 1,	14.71	11.62	8.89	8.84	7.85	7.63
1, 1, 1, 1, 1)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
(10, 10, 10, 10, 10,	14.71	11.62	9.54	9.51	9.07	9.00
1, 1, 1, 1, 1)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(25, 25, 25, 25, 25,	14.71	11.62	9.57	9.56	9.72	9.69
25, 25, 25, 1, 1)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(30, 30, 30, 1, 1,	14.71	11.62	9.74	9.72	9.27	9.20
1, 1, 1, 1, 1)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(50, 1, 1, 1, 1,	14.71	11.62	9.44	9.41	8.83	8.72
1, 1, 1, 1, 1)	(0.09)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)
(20, 20, 20, 5, 5,	14.71	11.62	9.58	9.56	9.27	9.24
5, 5, 1, 1, 1)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(512, 256, 128, 64, 32,	14.71	11.62	10.17	10.16	10.42	10.43
16, 8, 4, 2, 1)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(0.50, 0.45, 0.40, 0.35, 0.30,	14.71	11.62	9.19	9.16	8.80	8.71
0.25, 0.20, 0.15, 0.10, 0.05)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(10, 0.1, 0.1, 0.1, 0.1,	14.71	11.62	9.48	9.47	9.55	9.49
0.1, 0.1, 0.1, 0.1, 0.1)	(0.09)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)
(10, 5, 1, 0.1, 0.1,	14.71	11.62	9.85	9.84	9.72	9.70
0.1, 0.1, 0.1, 0.1, 0.1)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(10, 10, 10, 0.1, 0.1,	14.71	11.62	9.87	9.85	9.65	9.61
0.1, 0.1, 0.1, 0.1, 0.1)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(10, 10, 10, 10, 1,	14.71	11.62	10.07	10.06	9.88	9.88
1, 1, 0.1, 0.1, 0.1)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(10, 10, 5, 5, 1,	14.71	11.62	9.95	9.93	9.73	9.73
1, 0.4, 0.4, 0.1, 0.1)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(20, 6, 8/3, 14/9, 1,	14.71	11.62	9.94	9.92	9.75	9.77
2/3, 4/9, 2/7, 1/6, 2/27)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
(81, 27, 9, 3, 1,	14.71	11.62	10.43	10.42	10.44	10.48
1/2, 1/4, 1/8, 1/16, 1/32)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
$(10^3, 10^2, 25, 5, 2,$	14.71	11.62	11.02	11.02	11.23	11.25
$1/2, 1/5, 1/20, 10^{-2}, 10^{-3})$	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)

stantial risk reduction over $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$, but in showing that $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$ is inadmissible.

4. For the estimation of Σ_i under Stein's loss, the study indicates that $\hat{\Sigma}_i^{ST}$ and $\hat{\Sigma}_i^{HF}$ are close to being minimax.

Acknowledgments. This paper forms a part of my Ph.D. dissertation at Stanford University. I would like to express my deep gratitude to my advisor, Professor Charles Stein, whose constant support, encouragement and guidance made this dissertation possible. I would also like to thank Professors Dipak

Dey and L. R. Haff for sending me copies of their unpublished manuscripts. Finally I would like to thank an Associate Editor and two anonymous referees for detailed suggestions which considerably improved the quality of the paper.

REFERENCES

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

DEY, D. K. and SRINIVASAN, C. (1986). Trimmed minimax estimator of a covariance matrix. *Ann. Inst. Statist. Math.* **38** 101–108.

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

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

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

HAFF, L. R. (1988). The variational form of certain Bayes estimators. Unpublished manuscript.

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

LOH, W. L. (1988). Estimating covariance matrices. Ph.D. dissertation, Dept. Statist., Stanford Univ.

STEIN, C. (1956). Some problems in multivariate analysis. Part I. Technical Report 6, Dept. Statist., Stanford Univ.

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

STEIN, C. (1975). Estimation of a covariance matrix Rietz Lecture, 39th Annual IMS Meeting, Atlanta, Georgia. Unpublished manuscript.

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

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907