The Canadian Journal of Statistics
Vol. 17, No. 1, 1989, Pages 91–102
La Revue Canadienne de Statistique

Improved minimax estimation of a normal precision matrix

K. KRISHNAMOORTHY and A.K. GUPTA

Bowling Green State University

Key words and phrases: Best invariant estimators, Wishart distribution, fully invariant loss function, risk.
AMS 1985 subject classifications: Primary 62H12; secondary 62F10.

ABSTRACT

Let $S_{p\times p}$ have a Wishart distribution with parameter matrix Σ and n degrees of freedom. We consider here the problem of estimating the precision matrix Σ^{-1} under the loss functions $L_1(\Sigma^{-1}, \hat{\Sigma}^{-1}) = \operatorname{tr}(\hat{\Sigma}^{-1}\Sigma) - \log|\hat{\Sigma}^{-1}\Sigma| - p$ and $L_2(\Sigma^{-1}, \hat{\Sigma}^{-1}) = \operatorname{tr}(\hat{\Sigma}^{-1}\Sigma - I)^2$. James-Stein-type estimators have been derived for an arbitrary p. We also obtain an orthogonal invariant and a diagonal invariant minimax estimator under both loss functions. A Monte-Carlo simulation study indicates that the risk improvement of the orthogonal invariant estimators over the James-Stein type estimators, the Haff (1979) estimator, and the "testimator" given by Sinha and Ghosh (1987) is substantial.

RÉSUMÉ

Soit $S_{p\times p}$ une matrice aléatoire suivant une loi de Wishart avec n degrés de liberté et matrice des paramètres Σ. On étudie le problème de l'estimation de Σ^{-1} lorsqu'un utilise les fonctions de perte: $L_1(\Sigma^{-1}, \hat{\Sigma}^{-1}) = \operatorname{tr}(\hat{\Sigma}^{-1}\Sigma) - \log|\hat{\Sigma}^{-1}\Sigma| - p$ et $L_2(\Sigma^{-1}, \hat{\Sigma}^{-1}) = \operatorname{tr}(\hat{\Sigma}^{-1}\Sigma - I)^2$. Des estimateurs de type James-Stein sont définis pour n'importe quelle valeur de p. Relativement aux deux fonctions de perte considérées, on obtient des estimateurs minimax qui sont invariants sous les transformations orthogonales ou celles dont la matrice est triangulaire (inférieure). Une étude de simulation Monte-Carlo indique qu'on obtient une diminution importante du risque en utilisant les estimateurs invariants sous les transformations orthogonales plutôt que les estimateurs de type James-Stein, ou de Haff (1979), ou le "test-estimateur" proposé par Sinha et Ghosh (1987).

1. INTRODUCTION

Let $S \sim W_p(n, \Sigma)$, where Σ is unknown and $n > p + 1$. We are interested in estimating Σ^{-1} under the loss functions

$$L_1(\hat{\Sigma}^{-1}, \Sigma^{-1}) = \operatorname{tr}(\hat{\Sigma}^{-1}\Sigma) - \log|\hat{\Sigma}^{-1}\Sigma| - p \tag{1.1}$$

and

$$L_2(\hat{\Sigma}^{-1}, \Sigma^{-1}) = \operatorname{tr}(\hat{\Sigma}^{-1}\Sigma - I)^2. \tag{1.2}$$

The loss function L_1 is analogous to the one considered by James and Stein (1960) for estimating Σ, and L_2 is the multivariate generalization of the univariate quadratic loss function. Note that both loss functions are fully invariant and strictly convex.

The usual estimator of Σ^{-1} is $(n-p-1)S^{-1}$ and is the best multiple of S^{-1} under L_1. For the loss L_2, Haff (1979) has shown that the best multiple of S^{-1} is $\{(n-p-3)(n-p)/(n-1)\}S^{-1}$, and is dominated by

$$\hat{\Sigma}_{H_2}^{-1} = \frac{(n-p-3)(n-p)}{(n-1)}\,[S^{-1} + \upsilon t(\upsilon) I] \tag{1.3}$$

where $\upsilon = (\operatorname{tr} S)^{-1}$ and $t(\upsilon)$ is an absolutely continuous function such that $0 \leq t(\upsilon) \leq 2(p-1)/(n-p)$ and $t'(\upsilon) \leq 0$. No Haff-type estimator of Σ^{-1} is available for the loss L_1.

Olkin and Selliah (1977) (under L_1) and Sharma and Krishnamoorthy (1983) (under L_1 and L_2) have derived the best lower triangular invariant estimators of Σ^{-1} for the case $p=2$.

In Section 2 we derive, for any p, the best lower triangular invariant estimator, $\hat{\Sigma}_{li}^{-1}$, under the loss L_i $(i=1,2)$. In Section 3 we obtain a diagonal invariant estimator $\hat{\Sigma}_{di}^{-1}$ which dominates $\hat{\Sigma}_{li}^{-1}$ $(i = 1, 2)$.

Sharma and Krishnamoorthy (1983) derived an orthogonal invariant minimax estimator $\psi_i(s)$ of Σ^{-1} for $p = 2$, which is better than $\hat{\Sigma}_{li}^{-1}$, under the loss L_i $(i = 1, 2)$. Such orthogonal invariant estimators are not available for $p \geq 3$. In Section 4 we develop an orthogonal invariant estimator of Σ^{-1} for an arbitrary p, which is analogous to the estimator of Σ given by Dey and Srinivasan (1985).

Finally, in Section 5, we carry out a Monte-Carlo simulation study to indicate the nature of the risk improvement. The study indicates that our orthogonal invariant estimators are substantially better than Haff's (1979) estimator (under L_2), than Sinha and Ghosh's (1986) "testimator" (under L_1), and than the best lower triangular invariant estimators $\hat{\Sigma}^{-1}$ (under L_1 and L_2). Note that a "testimator" is not available for the loss L_2.

2. DERIVATION OF $\hat{\Sigma}_{li}^{-1}$ UNDER THE LOSS L_i $(i = 1, 2)$

Let $S \sim W_p(n, \Sigma)$, and consider the group G_A of lower triangular matrices $\boldsymbol{A}$ acting on $\{\boldsymbol{S}:\boldsymbol{S}>0\}$ as $\boldsymbol{S} \to \boldsymbol{ASA}'$. This induces the transformation $\Sigma^{-1} \to \boldsymbol{A}'^{-1}\Sigma^{-1}\boldsymbol{A}^{-1}$, $\hat{\Sigma}^{-1}(\boldsymbol{S}) \to \boldsymbol{A}'^{-1}\hat{\Sigma}^{-1}(\boldsymbol{S})\boldsymbol{A}^{-1}$, and with the loss functions L_1 and L_2, the estimation problem remains invariant under G_A. It is easy to verify that any estimator invariant under G_A must be of the form $\boldsymbol{T}'^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}$, where $\boldsymbol{T}$ is a lower triangular matrix with positive diagonal elements such that $\boldsymbol{TT}' = \boldsymbol{S}$, and $\boldsymbol{D}$ is a diagonal matrix whose elements are independent of $\boldsymbol{S}$. Since the group G_A acts transitively on the parameter space $\{\Sigma : \Sigma > 0\}$, the risk function $R_i(\Sigma^{-1}, \boldsymbol{T}'^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}) = \mathscr{E}L_i(\Sigma^{-1}, \boldsymbol{T}'^{-1}\boldsymbol{D}\boldsymbol{T}^{-1})$ $(i = 1, 2)$ is independent of Σ.

In the following theorem we derive the best lower triangular invariant estimator $\hat{\Sigma}_{li}^{-1}$ under L_1.

THEOREM 2.1. *For the loss L_1, the best lower triangular invariant estimator is given by*

$$\hat{\Sigma}_{l1}^{-1} = \boldsymbol{T}'^{-1}\Delta_1\boldsymbol{T}^{-1} \tag{2.1}$$

where $\Delta_1 = \operatorname{diag}(\Delta_{11}, \Delta_{12}, \ldots, \Delta_{1p})$ *and* $\Delta_{1i} = (n-i-1)(n-i)/(n-1)$, $i = 1, 2, \ldots, p$.

Proof. Since the risk $R_1(\Sigma^{-1}, \boldsymbol{T}'^{-1}\boldsymbol{D}\boldsymbol{T}^{-1})$ is independent of Σ, without loss of generality, we may take $\Sigma = \boldsymbol{I}$. Then

$$\begin{aligned} R_1(\boldsymbol{I}, \boldsymbol{T}'^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}) &= \mathscr{E}\operatorname{tr}(\boldsymbol{T}'^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}) - \mathscr{E}\log|\boldsymbol{T}'^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}| - p \\ &= \mathscr{E}\operatorname{tr}(\boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{D} + \mathscr{E}\log|\boldsymbol{S}| - \sum_1^p \log d_i - p. \end{aligned}$$

From Section AI of the Appendix we have $\mathscr{E}\operatorname{tr}(\boldsymbol{T}'\boldsymbol{T})^{-1} = \operatorname{tr}\boldsymbol{H} = \sum_1^p h_i$, where

$$h_i = \frac{n-1}{(n-i)(n-i-1)}, \qquad i = 1, 2, \ldots, p. \tag{2.2}$$

We can rewrite

$$R_1(I, T'^{-1}DT^{-1}) = \sum_1^p h_i d_i - \sum_1^p \log d_i + \mathscr{E} \log|S| - p,$$

which is minimum when $d_i = h_i^{-1} = \Delta_{1i}$, $i = 1, 2, \ldots, p$. Q.E.D.

We next derive the best lower triangular invariant estimator $\hat{\Sigma}_{l2}^{-1}$ under the loss L_2.

THEOREM 2.2. *For the loss L_2, the best lower triangular invariant estimator is given by*

$$\hat{\Sigma}_{l2}^{-1} = T'^{-1}\Delta_2 T^{-1}, \tag{2.3}$$

where $\Delta_2 = \operatorname{diag}(\Delta_{21}, \Delta_{22}, \ldots, \Delta_{2p})$, $\Delta_{2i} = d_i$ $(i = 1, 2, \ldots, p)$ is the solution of the equations

$$q_j = \frac{n-1}{(n-j)(n-j-1)}, \qquad j = 1, 2, \ldots, p, \tag{2.4}$$

and the q_j's are given by (A.8) *of the Appendix.*

Proof. For the reason given in Theorem 2.1, without loss of generality we take $\Sigma = I$. We have

$$\begin{aligned} R_2(I, T'^{-1}DT^{-1}) &= \mathscr{E}\operatorname{tr}(T'^{-1}DT^{-1} - I)^2 \\ &= \mathscr{E}\operatorname{tr}(T'T)^{-1}D(T'T)^{-1}D - 2\mathscr{E}\operatorname{tr}(T'T)^{-1}D + p. \end{aligned} \tag{2.5}$$

From the results reported in the Appendix we get $\mathscr{E}(T'T)^{-1} = \operatorname{diag}(h_1, h_2, \ldots, h_p)$ and $\mathscr{E}(T'T)^{-1}D(T'T)^{-1} = \operatorname{diag}(q_1, \ldots, q_p)$, where the h_j's are given in (2.2) and the q_j's in (A.8). Substituting the above relations in (2.5), we get

$$R_2(I, T'^{-1}DT^{-1}) = \sum_{j=1}^p q_j d_j - 2\sum_{j=1}^p h_j d_j + p.$$

Differentiating the above equation with respect to the d_j's and equating the derivatives to zero, we obtain

$$q_j + \sum_{j=1}^p d_i \frac{\partial q_i}{\partial d_j} = \frac{2(n-1)}{(n-j)(n-j-1)}, \qquad j = 1, 2, \ldots, p. \tag{2.6}$$

Since $(T'T)^{-1}$ is symmetric and $\mathscr{E}(T'T)^{-1}D(T'T)^{-1}$ is a diagonal matrix, we can write the diagonal entries q_j as

$$(q_1, q_2, \ldots, q_p)' = A(d_1, d_2, \ldots, d_p)',$$

where A is a symmetric matrix whose (i,j)th element is the expected value of the square of (i,j)th element of $(T'T)^{-1}$. This implies that

$$\sum_{i=1}^p d_i \frac{\partial q_i}{\partial d_j} = q_j \qquad (j = 1, 2, \ldots, p),$$

REMARK. For $p = 2$ and 3 one can find the solution of the equation (2.4). For $p \geq 4$, solving (2.4) will be quite messy. Here we give the solution for $p = 2$ and 3. For a given

$p \geq 4$, one can use the FORTRAN subroutine called REDUCE to get the analytic solution. For $p = 2$,

$$\Delta_{21} = \frac{(n-1)(n-4)\{(n-3)^2 - (n-5)\}}{(n-1)(n-3)^2 - (n-5)},$$
$$\Delta_{22} = \frac{(n-2)(n-3)(n-4)(n-5)}{(n-1)(n-3)^2 - (n-5)}, \tag{2.7}$$

and for $p = 3$,

$$\Delta_{21} = (n-4) - \frac{\Delta_{22}}{n-3} - \frac{\Delta_{23}(n-2)}{(n-4)(n-3)},$$
$$\Delta_{22} = \frac{(n-2)(n-3)(n-4)(n-5)}{(n-1)(n-3)^2 - (n-5)} - \frac{(n-2)(n-3)(n-5)(n-6)}{(n-1)[(n-2)(n-3)(n-4) - 2(n-6)]},$$
$$\Delta_{23} = \frac{(n-2)(n-3)(n-4)(n-5)(n-6)}{(n-1)[(n-2)(n-3)(n-4) - 2(n-6)]}. \tag{2.8}$$

Since the risk of $T'^{-1}DT^{-1}$ is constant under L_i, the best invariant estimator $\hat{\Sigma}_{li}^{-1}$ $(i = 1, 2)$ is minimax among the estimators invariant under G_A. Also, as G_A is solvable, from Kiefer's theorem (Kiefer 1957), $\hat{\Sigma}_{li}^{-1}$ $(i = 1, 2)$ is minimax. For any $b > 0$, bS^{-1} is invariant under G_A and, being different from $\hat{\Sigma}_{li}^{-1}$ $(i = 1, 2)$, it is inadmissible.

For the loss L_1, the minimax risk is given by

$$\begin{aligned} R_1(I, \hat{\Sigma}_{l1}^{-1}) &= \mathscr{E} \log |S| - \sum_1^p \log \Delta_{1i} \\ &= \sum_1^p \mathscr{E} \log \chi^2_{n-j+1} - \sum_1^p \log \Delta_{1i} \\ &= p \log 2 + \sum_1^p \psi\left(\frac{n-j+1}{2}\right) - \sum_1^p \log \Delta_{1i}, \end{aligned} \tag{2.9}$$

where $\psi(x)$ is the digamma function and the Δ_{1i}'s are given in Theorem 2.1.

For the loss L_2, since $R_2(\boldsymbol{I}, \boldsymbol{T}'^{-1}\boldsymbol{D}\boldsymbol{T}^{-1})$ is minimized when $\boldsymbol{D} = \boldsymbol{\Delta}_2$, we have

$$\mathscr{E}\, \mathrm{tr}\,(\boldsymbol{T}'^{-1}\boldsymbol{\Delta}_2\boldsymbol{T}^{-1}\boldsymbol{T}'^{-1}\boldsymbol{\Delta}_2\boldsymbol{T}^{-1}) = \mathscr{E}\, \mathrm{tr}\,(\boldsymbol{T}'^{-1}\boldsymbol{\Delta}_2\boldsymbol{T}^{-1})$$

and the minimax risk is given by

$$\begin{aligned} R_2(\boldsymbol{I}, \boldsymbol{T}'^{-1}\boldsymbol{\Delta}_2\boldsymbol{T}^{-1}) &= p - \mathscr{E}\, \mathrm{tr}\,(\boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{\Delta}_2 \\ &= p - \sum_1^p \Delta_{2j} h_j. \end{aligned} \tag{2.10}$$

3. DERIVATION OF $\hat{\Sigma}_{di}^{-1}$ UNDER L_i $(i = 1, 2)$

Let $L(\Sigma^{-1}, \hat{\Sigma}^{-1})$ be a fully invariant loss function. If $\phi(S)$ is a constant-risk minimax estimator, so is $\phi_B(S) = \boldsymbol{B}\phi(\boldsymbol{B}'\boldsymbol{S}\boldsymbol{B})\boldsymbol{B}'$ for any nonsingular matrix $\boldsymbol{B}$ (e.g. see Sharma and

Krishnamoorthy 1983). In particular, letting $\phi_i(S) = \hat{\Sigma}_{li}^{-1}$, we have the family

$$\{\phi_{i\Gamma}(S) = \Gamma\phi(\Gamma' S\Gamma)\Gamma' : \Gamma \in G_\Gamma, \text{ the group of orthogonal matrices}\} \tag{3.1}$$

of constant-risk minimax estimators under L_i ($i = 1, 2$). Note that when $\Gamma = I$, $\phi_{i\Gamma}(S) = \hat{\Sigma}_{li}^{-1}$, and for $\Gamma = (\gamma_{ij})$ with $\gamma_{ij} = 1$ if $i + j = p + 1$ and 0 otherwise,

$$\phi_{i\Gamma}(S) = \hat{\Sigma}_{ui}^{-1} = U'^{-1}\Delta_i^0 U^{-1} \tag{3.2}$$

where U is an upper triangular matrix with positive diagonal elements such that $UU' = S$, and $\Delta_i^0 = \operatorname{diag}(\Delta_{ip}, \Delta_{ip-1}, \ldots, \Delta_{i1})$ is the best upper triangular invariant estimator of Σ^{-1} under L_i ($i = 1, 2$).

As L_i is a strictly convex function of $\hat{\Sigma}^{-1}$, any average $\alpha\phi_{i\Gamma}(S) + (1 - \alpha)\phi_{i\eta}(S)$, $\eta \neq \Gamma$, is better than $\hat{\Sigma}_{li}^{-1}$ ($i = 1, 2$) and so it is minimax. In the following theorem we prove that $\frac{1}{2}$ is the best choice of α.

THEOREM 3.1. *The best choice of* α *which minimizes the risk of*

$$\hat{\Sigma}_\alpha^{-1} = \alpha\phi_{2\Gamma}(S) + (1 - \alpha)\phi_{2\eta}(S)$$

under the loss L_2 *is* $\frac{1}{2}$.

Proof. We first observe that

$$R_2(\Sigma^{-1}, \phi_{2\Gamma}(S)) = R_2(\Sigma^{-1}, \phi_{2\eta}(S)) \quad \text{and} \quad \mathscr{E} \operatorname{tr} \phi_{2\Gamma}(S)\Sigma^{-1} = \mathscr{E} \operatorname{tr} \phi_{2\eta}(S)\Sigma^{-1}. \tag{3.3}$$

Let $\hat{\Sigma}_{1-\alpha}^{-1} = (1 - \alpha)\phi_{2\Gamma}(S) + \alpha\phi_{2\eta}(S)$. Then the equations in (3.3) imply that $R_2(\Sigma^{-1}, \hat{\Sigma}_\alpha^{1}) = R_2(\Sigma^{-1}, \hat{\Sigma}_{1-\alpha}^{-1})$ for any α, $0 < \alpha < 1$. Suppose that some $\alpha \neq \frac{1}{2}$ is the best choice. Then, for any β, $0 < \beta < 1$, $\beta\hat{\Sigma}_\alpha^{-1} + (1 - \beta)\hat{\Sigma}_{1-\alpha}^{-1}$ is different from $\hat{\Sigma}_\alpha^{-1}$ and $\hat{\Sigma}_{1-\alpha}^{-1}$, and dominates each of them. Thus, we arrive at a contradiction. Q.E.D.

For the loss L_1, we do not know the best choice of α for which $R_1(\Sigma^{-1}, \alpha\phi_{1\Gamma}(S) + (1 - \alpha)\phi_{1\eta}(S))$ is minimized. A sufficient condition for $\frac{1}{2}$ to be the best choice of α is

$$R_1(\Sigma^{-1}, \alpha\phi_{1\Gamma}(S) + (1 - \alpha)\phi_{1\eta}(S)) = R_1(\Sigma^{-1}, (1 - \alpha)\phi_{1\Gamma}(S) + \alpha\phi_{1\eta}(S)).$$

Although one can consider a simple average of any two members of the family (3.1), we take

$$\hat{\Sigma}_{di}^{-1} = \frac{\hat{\Sigma}_{li}^{-1} + \hat{\Sigma}_{ui}^{-1}}{2}, \tag{3.4}$$

since $\hat{\Sigma}_{li}^{-1}$ and $\hat{\Sigma}_{ui}^{-1}$ are easy to compute. Note that $\hat{\Sigma}_{di}^{-1}$ is diagonal invariant.

4. DERIVATION OF $\hat{\Sigma}_{0i}^{-1}$ ($i = 1, 2$)

Though the best lower triangular invariant estimator $\hat{\Sigma}_{li}^{-1}$ ($i = 1, 2$) is minimax, it suffers from the fact that some elements of Σ^{-1} are grossly underestimated while other elements are overestimated. As pointed out by Eaton (1970), this is because "the standard orthonormal basis in $\mathbb{R}^p$ plays a vital role under the action of elements of G_A". A way out is to give equal importance to all orthonormal coordinate systems. In other words, we will look for an orthogonal invariant minimax estimator.

Sharma and Krishnamoorthy (1983) derived an orthogonal invariant minimax estimator

$$\psi_i(S) = \int_{G_\Gamma} \phi_{i\Gamma}(S)\, d\upsilon(\Gamma) \tag{4.1}$$

TABLE 1: Risks of b_iS^{-1} and $\hat{\Sigma}_{li}^{-1}$ under L_i, $i = 1, 2$.[a]

	$p = 2$		$p = 3$	
	$n = 10$	$n = 20$	$n = 10$	$n = 20$
$R_1(b_1S^{-1})$	0.3837	0.1694	0.8553	0.3518
$R_1(\hat{\Sigma}_{l1}^{-1})$	0.3737	0.1663	0.7825	0.3387
$R_2(b_2S^{-1})$	0.7302	0.3282	1.4444	0.6513
$R_2(\hat{\Sigma}_{l2}^{-1})$	0.6995	0.3222	1.3201	0.6274

[a] $b_1 = (n - p - 1)$, $b_2 = (n - p - 3)(n - p)/(n - 1)$.

of Σ^{-1} under L_i $(i = 1, 2)$ for the case $p = 2$, where $\phi_{i\Gamma}(S)$ is given in (3.1) and υ is an invariant Haar measure over the orthogonal group G_Γ. For $p \geq 3$, evaluation of the integral (4.1) seems to be difficult and $\psi_i(S)$ is not available explicitly.

We develop here an orthogonal invariant estimator of Σ^{-1} for an arbitrary p. Note that

$$\Phi_{i\Gamma}(S) = \Gamma\phi_i(\Gamma' S\Gamma)\Gamma' \tag{4.2}$$

is a constant-risk minimax estimator for any orthogonal matrix Γ independent of S. Let S have the spectral decomposition $S = RLR'$ where $RR' = I$, $L = \operatorname{diag}(l_1, \ldots, l_p)$, and $l_1 > l_2 > \cdots > l_p > 0$. In (4.2), if we let $\Gamma = R$, then

$$\hat{\Sigma}_{0i}^{-1} = \phi_{iR}(S) = R\phi_i(L)R' \tag{4.3}$$

becomes an orthogonal invariant estimator with

$$\phi_i(L) = \operatorname{diag}(\Delta_{i1} l_1^{-1}, \Delta_{i2} l_2^{-1}, \ldots, \Delta_{ip} l_p^{-1}) \qquad (i = 1, 2).$$

The estimator (4.3) is analogous to the estimator of Σ given by Dey and Srinivasan (1985). They have shown that, under $L(\Sigma, \hat{\Sigma}) = \operatorname{tr}(\Sigma\hat{\Sigma}^{-1}) - \log|\hat{\Sigma}\Sigma^{-1}| - p$, their orthogonal invariant estimator is minimax using the "unbiased estimator of the risk expression of any orthogonal invariant estimator" given by Haff (1982). Deriving such an unbiased estimate of the risk in the present problem seems to be difficult, and we do not know, theoretically, if the estimator $\hat{\Sigma}_{0i}^{-1}$ dominates $\hat{\Sigma}_{li}^{-1}$ $(i = 1, 2)$. However, the Monte Carlo simulation study in Section 5 indicates that $\hat{\Sigma}_{0i}^{-1}$ is not only minimax but is also substantially better than $\hat{\Sigma}_{li}^{-1}$ $(i = 1, 2)$. We also observe that the estimator $\hat{\Sigma}_{di}^{-1}$ shrinks the larger eigenvalues and expands the smaller eigenvalues toward some central value. All these lead us to make the following conjecture.

CONJECTURE 4.1. Let $L(\Sigma^{-1}, \hat{\Sigma}^{-1})$ be a fully invariant and strictly convex loss function. If $\phi(S)$ is a constant-risk minimax estimator of Σ^{-1}, then $R\phi(L)R'$, where $RLR' = S$, $RR' = I$, and $L = \operatorname{diag}(l_1, l_2, \ldots, l_p)$ with $l_1 > l_2 > \cdots > l_p > 0$, is orthogonal invariant and better than $\phi(S)$. The same type of conjecture may also be stated for the estimation of Σ.

5. MONTE CARLO SIMULATION STUDY

In this section we compute the risks of Haff's estimator (1.3) and of the estimators presented in this paper, and compare their performances. For convenience we shall write $R_i(\Sigma^{-1}, \hat{\Sigma}_i^{-1}) = R_i(\hat{\Sigma}_i^{-1})$.

Table 1 gives the exact risks of $\hat{\Sigma}_{li}^{-1}$, of the best lower triangular invariant minimax estimator, and of b_iS^{-1}, the best multiple of S^{-1} under the loss L_i $(i = 1, 2)$. The risks of the estimators $\hat{\Sigma}_{di}^{-1}$, $\hat{\Sigma}_{0i}^{-1}$, and $\hat{\Sigma}_{H2}^{-1}$ have been estimated from the generation of 2000 inde-

TABLE 2: Risks of $\hat{\Sigma}_{di}^{-1}$ and $\hat{\Sigma}_{0i}^{-1}$ under L_i ($i = 1, 2$) for $p = 2$.

	$n = 10$		$n = 20$		$n = 10$		$n = 20$	
	$R_1(\hat{\Sigma}_{d1}^{-1})$	$R_1(\hat{\Sigma}_{01}^{-1})$	$R_1(\hat{\Sigma}_{d1}^{-1})$	$R_1(\hat{\Sigma}_{01}^{-1})$	$R_2(\hat{\Sigma}_{d2}^{-1})$	$R_2(\hat{\Sigma}_{02}^{-1})$	$R_2(\hat{\Sigma}_{d2}^{-1})$	$R_2(\hat{\Sigma}_{02}^{-1})$
$\Sigma = (\sigma_{ij})$; $\sigma_{ii} = 1$, $i = 1, 2$, and $\sigma_{12} = \rho$								
$\rho =$								
0	0.363 (0.006)	0.295 (0.005)	0.164 (0.002)	0.140 (0.001)	0.664 (0.015)	0.585 (0.012)	0.315 (0.005)	0.270 (0.005)
0.10	0.365 (0.006)	0.303 (0.005)	0.165 (0.002)	0.142 (0.001)	0.667 (0.009)	0.601 (0.010)	0.317 (0.005)	0.279 (0.005)
0.30	0.367 (0.005)	0.317 (0.004)	0.167 (0.003)	0.145 (0.003)	0.669 (0.010)	0.608 (0.013)	0.318 (0.005)	0.285 (0.006)
0.50	0.368 (0.005)	0.338 (0.005)	0.168 (0.002)	0.152 (0.003)	0.670 (0.008)	0.612 (0.009)	0.319 (0.005)	0.289 (0.006)
0.80	0.373 (0.005)	0.370 (0.005)	0.169 (0.003)	0.165 (0.002)	0.680 (0.009)	0.660 (0.009)	0.321 (0.005)	0.310 (0.006)
0.99	0.380 (0.005)	0.382 (0.005)	0.170 (0.004)	0.169 (0.003)	0.692 (0.014)	0.692 (0.015)	0.323 (0.005)	0.321 (0.005)
$\Sigma = \text{diag}(1, c)$								
$c =$								
0.01	0.364 (0.008)	0.370 (0.008)	0.168 (0.002)	0.168 (0.003)				
0.10	0.364 (0.008)	0.363 (0.007)	0.168 (0.002)	0.166 (0.003)				
0.30	0.364 (0.008)	0.335 (0.007)	0.168 (0.002)	0.160 (0.002)				
0.50	0.364 (0.008)	0.313 (0.006)	0.168 (0.002)	0.151 (0.003)				
0.80	0.364 (0.008)	0.297 (0.007)	0.168 (0.002)	0.142 (0.002)				
1.0	0.364 (0.008)	0.295 (0.007)	0.168 (0.002)	0.140 (0.002)				

TABLE 3: Risks of $\hat{\Sigma}_{H2}^{-1}$ and $\hat{\Sigma}_{02}^{-1}$ under L_2.[a]

	$n = 10$		$n = 20$	
	$R_2(\hat{\Sigma}_{H2}^{-1})$	$R_2(\hat{\Sigma}_{02}^{-1})$	$R_2(\hat{\Sigma}_{H2}^{-1})$	$R_2(\hat{\Sigma}_{02}^{-1})$
	$\Sigma = \text{diag}(1, c)$, $p = 2$			
$c =$ 0.01	0.703 (0.011)	0.672 (0.009)	0.318 (0.007)	0.301 (0.007)
0.1	0.699 (0.012)	0.635 (0.008)	0.316 (0.006)	0.300 (0.006)
0.3	0.695 (0.009)	0.590 (0.008)	0.313 (0.006)	0.298 (0.006)
0.5	0.689 (0.008)	0.588 (0.007)	0.313 (0.006)	0.292 (0.007)
0.7	0.686 (0.009)	0.581 (0.006)	0.312 (0.004)	0.288 (0.006)
0.9	0.684 (0.008)	0.580 (0.006)	0.310 (0.005)	0.279 (0.006)
1.0	0.680 (0.007)	0.578 (0.009)	0.310 (0.005)	0.274 (0.005)
	$\Sigma = \text{diag}(1, c_1, c_2)$, $p = 3$			
$(c_1, c_2) =$ (0.01, 0.01)	1.39 (0.012)	1.24 (0.009)	0.639 (0.009)	0.613 (0.008)
(0.1, 0.1)	1.38 (0.011)	1.17 (0.011)	0.634 (0.010)	0.601 (0.007)
(0.1, 0.2)	1.37 (0.013)	1.14 (0.009)	0.631 (0.008)	0.599 (0.008)
(0.2, 0.5)	1.37 (0.011)	1.06 (0.011)	0.630 (0.007)	0.561 (0.008)
(0.3, 0.6)	1.36 (0.010)	1.04 (0.014)	0.625 (0.009)	0.522 (0.008)
(0.4, 0.7)	1.37 (0.009)	1.03 (0.008)	0.621 (0.009)	0.501 (0.008)
(0.8, 0.9)	1.37 (0.009)	1.02 (0.009)	0.618 (0.008)	0.467 (0.008)
(1, 1)	1.36 (0.008)	1.02 (0.009)	0.618 (0.008)	0.457 (0.01)

[a] The numbers in parentheses represent the estimated values of the standard error.

pendent samples from a $W_p(n, \Sigma)$. The FORTRAN subroutine given by Smith and Hocking (1972) was used. For each risk estimation we also computed the sample standard error.

We present the values of $R_i(\hat{\Sigma}_{0i}^{-1})$ and $R_i(\hat{\Sigma}_{di}^{-1})$ $(i = 1, 2)$ in Tables 2 (for $p = 2$) and 4 (for $p = 3$). As $\hat{\Sigma}_{di}^{-1}$ is diagonal invariant and $\hat{\Sigma}_{0i}^{-1}$ is orthogonal invariant, for the purpose of comparing them, their risks are computed only for the cases where Σ is a diagonal or a correlation matrix. We also note that $R_i(\Sigma, \hat{\Sigma}_{di}^{-1}) = R_i(I, \hat{\Sigma}_{di}^{-1})$ $(i = 1, 2)$ for all $\Sigma = \text{diag}(c_1, c_2, \ldots, c_p)$. Tables 2 and 4 indicate that $\hat{\Sigma}_{0i}^{-1}$ performs better than $\hat{\Sigma}_{di}^{-1}$ under the loss L_i $(i = 1, 2)$ except when $|\Sigma|$ is near zero. We also observe that $\hat{\Sigma}_{0i}^{-1}$ is not only minimax but also dominates the minimax estimator $\hat{\Sigma}_{li}^{-1}$ $(i = 1, 2)$ substantially for smaller as well as for larger values of n.

We compare the estimator $\hat{\Sigma}_{02}^{-1}$ with $\hat{\Sigma}_{H2}^{-1}$ in Table 3. To compute the risk of $\hat{\Sigma}_{H2}^{-1}$, we take $t(\upsilon) = (p-1)/(n-p)$ in (1.3). Since both $\hat{\Sigma}_{02}^{-1}$ and $\hat{\Sigma}_{H2}^{-1}$ are scale and orthogonal invariant, to compute their risks we can take $\Sigma = \text{diag}(1, c_1 c_2, \ldots, c_{p-1})$. Again, Table 3 indicates that $\hat{\Sigma}_{02}^{-1}$ dominates $\hat{\Sigma}_{H2}^{-1}$ uniformly.

Comparison of $R_i(\hat{\Sigma}_{0i}^{-1})$ with $R_i(\psi_i(S))$ [$\psi_i(S)$ is given in (4.1)] computed in Sharma and Krishnamoorthy (1983) shows that $\hat{\Sigma}_{0i}^{-1}$ dominates $\psi_i(S)$ $(i = 1, 2)$ except when $|\Sigma|$ is near zero. We also compared $R_1(\hat{\Sigma}_{01}^{-1})$ with the risks of the "testimator" computed in Sinha and Ghosh (1987). We again infer that $\hat{\Sigma}_{01}^{-1}$ is uniformly substantially better than the "testimator".

APPENDIX

AI.

We need to compute $H = \mathscr{E}(T'T)^{-1}$, where $T = (t_{ij})$ is a lower triangular matrix with positive diagonal elements such that $TT' = S \sim W_p(n, I)$. We know that the t_{ij}'s $(i \geq j)$ are independent with $t_{ii}^2 \sim \chi^2_{n-i+1}$ $(i = 1, 2, \ldots, p)$ and $t_{ij} \sim \text{N}(0, 1)$ $(i \neq j)$. So for any

TABLE 4: Risks of $\hat{\Sigma}_{di}^{-1}$ and $\hat{\Sigma}_{0i}^{-1}$ under L_i ($i = 1, 2$) for $p = 3$.[a]

	$R_1(\hat{\Sigma}_{d1}^{-1})$	$R_1(\hat{\Sigma}_{01}^{-1})$	$R_1(\hat{\Sigma}_{d1}^{-1})$	$R_1(\hat{\Sigma}_{01}^{-1})$	$R_2(\hat{\Sigma}_{d2}^{-1})$	$R_2(\hat{\Sigma}_{02}^{-1})$	$R_2(\hat{\Sigma}_{d2}^{-1})$	$R_2(\hat{\Sigma}_{02}^{-1})$
$\Sigma = (\sigma_{ij})$; $\sigma_{ii} = 1$, $i = 1, 2, 3$; $\sigma_{12} = \rho_1$, $\sigma_{13} = \rho_2$, and $\sigma_{23} = \rho_3$								
$(\rho_1, \rho_2, \rho_3) =$								
(0, 0, 0)	0.692 (0.010)	0.532 (0.011)	0.306 (0.006)	0.252 (0.006)	1.159 (0.13)	1.024 (0.019)	0.596 (0.009)	0.456 (0.009)
(0.1, −0.1, 0.1)	0.693 (0.009)	0.531 (0.009)	0.306 (0.006)	0.256 (0.007)	1.17 (0.011)	1.03 (0.015)	0.598 (0.010)	0.465 (0.012)
(0.5, 0, 0)	0.696 (0.011)	0.589 (0.008)	0.306 (0.006)	0.258 (0.007)	1.16 (0.009)	1.04 (0.009)	0.600 (0.014)	0.511 (0.013)
(0.7, 0, 0)	0.700 (0.009)	0.643 (0.006)	0.309 (0.008)	0.280 (0.005)	1.17 (0.012)	1.06 (0.010)	0.602 (0.011)	0.550 (0.012)
(0, 0.85, 0)	0.736 (0.011)	0.689 (0.009)	0.312 (0.009)	0.300 (0.004)	1.43 (0.015)	1.25 (0.015)	0.602 (0.014)	0.540 (0.011)
(0.9, 0.01, −0.01)	0.705 (0.012)	0.722 (0.005)	0.315 (0.008)	0.309 (0.004)	1.20 (0.011)	1.13 (0.009)	0.607 (0.010)	0.579 (0.009)
(0.01, 0.01, 0.99)	0.701 (0.009)	0.719 (0.006)	0.307 (0.009)	0.310 (0.005)	1.23 (0.013)	1.20 (0.009)	0.612 (0.009)	0.581 (0.009)
(0.99, 0, 0)	0.710 (0.011)	0.721 (0.010)	0.309 (0.010)	0.316 (0.006)	1.20 (0.012)	1.17 (0.017)	0.610 (0.012)	0.590 (0.013)
(0.9, −0.8, −0.7)	0.720 (0.010)	0.739 (0.008)	0.310 (0.008)	0.320 (0.008)	1.26 (0.009)	1.23 (0.009)	0.640 (0.008)	0.637 (0.009)
$\Sigma = \mathrm{diag}(1, c_1, c_2)$								
$(c_1, c_2) =$								
(0.01, 0.01)	0.693 (0.009)	0.701 (0.009)	0.307 (0.006)	0.316 (0.007)				
(0.1, 0.2)	0.693 (0.009)	0.685 (0.017)	0.307 (0.006)	0.321 (0.007)				
(0.2, 0.5)	0.693 (0.009)	0.637 (0.015)	0.307 (0.006)	0.309 (0.006)				
(0.2, 0.9)	0.693 (0.009)	0.647 (0.016)	0.307 (0.006)	0.307 (0.007)				
(0.4, 0.7)	0.693 (0.009)	0.580 (0.017)	0.307 (0.006)	0.283 (0.006)				
(0.9, 0.9)	0.693 (0.009)	0.536 (0.016)	0.307 (0.006)	0.257 (0.006)				
(1, 1)	0.693 (0.009)	0.535 (0.015)	0.307 (0.006)	0.256 (0.006)				

[a]The numbers in parentheses represent the estimated values of the standard error.

diagonal matrix D with ± 1 on the diagonal, DTD and T have the same distribution, and

$$H = \mathscr{E}(T'T)^{-1} = \mathscr{E}[(DTD)'(DTD)]^{-1} = D\mathscr{E}(T'T)^{-1}D = DHD$$

implies that H must be a diagonal matrix. Using this fact, Eaton and Olkin (1987) obtained

$$h_j = \frac{n-1}{(n-j)(n-j-1)}, \qquad j = 1, 2, \ldots, p, \tag{A.1}$$

for the jth diagonal element of H.

AII.

To evaluate $Q = \mathscr{E}(T'T)^{-1}D(T'T)^{-1}$ we need the following lemmas.

LEMMA A.1. *Let $X \sim N_p(0, I)$ be independent of $S \sim W_p(n, I)$. Then*

$$X'(T'T)^{-1}X \sim \frac{p}{n-p+1} F_{p,n-p+1},$$

where $F_{p,n-p-1}$ is Snedecor's F-distribution with $(p, n-p-1)$ degrees of freedom and T is a lower triangular matrix with positive diagonal elements such that $TT' = S$.

See Theorem 2.2 of Tan and Guttman (1971).

LEMMA A.2. *Let $S \sim W_p(n, \Sigma)$. Then*

$$\mathscr{E}(S^{-2}) = \frac{\operatorname{tr}(\Sigma^{-1})\,\Sigma^{-1}}{(n-p)(n-p-1)(n-p-3)} + \frac{\Sigma^{-2}}{(n-p)(n-p-3)}.$$

For example, see Haff (1979).

Using the procedure given in Section AI, one can show that Q is a diagonal matrix. Let $\alpha_i \equiv \mathscr{E}(t_{ii})^{-2} = (n-i-1)^{-1}$ and $\beta_i \equiv \mathscr{E}(t_{ii})^{-4} = (n-i-1)^{-1}(n-i-3)^{-1}$, $i = 1, 2, \ldots, p$. Writing

$$S = \begin{pmatrix} S_{11\,(p-1)\times(p-1)} & S_{12} \\ S_{21} & S_{22\,(1\times 1)} \end{pmatrix} = TT'$$

and partitioning T accordingly, one can easily obtain

$$Q = \begin{bmatrix} Q_{11\,(p-1)\times(p-1)} & O \\ O & Q_{22\,(1\times 1)} \end{bmatrix},$$

where

$$\begin{aligned} Q_{11} &= \mathscr{E}\{(T'_{11}T_{11})^{-1}D_1(T'_{11}T_{11})^{-1} + (T'_{11}T_{11})^{-1}T'_{21}T'^{-1}_{22}D_2T^{-1}_{22}T_{21}(T'_{11}T_{11})^{-1}\} \\ &= \mathscr{E}\{(T'_{11}T_{11})^{-1}(D_1 + \alpha_p d_p I)(T'_{11}T_{11})^{-1}\}, \\ Q_{22} &= \mathscr{E}\{T^{-1}_{22}A_{21}T'^{-1}_{22}D_2T^{-1}_{22}A_{21}T'^{-1}_{22} + T^{-1}_{22}A_{21}T'^{-1}_{22}D_2(T'_{22}T_{22})^{-1} \\ &\quad + (T'_{22}T_{22})^{-1}D_2T^{-1}_{22}A_{21}T'^{-1}_{22} + (T'_{22}T_{22})^{-1}D_2(T'_{22}T_{22})^{-1} \\ &\quad + T^{-1}_{22}T_{21}(T'_{11}T_{11})^{-1}D_1(T'_{11}T_{11})^{-1}T'_{21}T'^{-1}_{22}\}, \\ A_{21} &= T_{21}(T'_{11}T_{11})^{-1}T'_{21}, \qquad D_1 = \operatorname{diag}(d_1, d_2, \ldots, d_{p-1}), \quad \text{and} \quad D_2 = d_p. \end{aligned}$$

Let $q_1, q_2, \ldots, q_p$ denote the diagonal elements of $\boldsymbol{Q}$. Then

$$q_p = d_p \mathscr{E}(t_{pp})^{-4}\{\mathscr{E}(\boldsymbol{A}_{21})^2 + 2\mathscr{E}(\boldsymbol{A}_{21}) + 1\} + \mathscr{E}(t_{pp})^{-2}\mathscr{E}\,\mathrm{tr}\{\boldsymbol{T}_{21}'\boldsymbol{T}_{21}(\boldsymbol{T}_{11}'\boldsymbol{T}_{11})^{-1}\boldsymbol{D}_1(\boldsymbol{T}_{11}'\boldsymbol{T}_{11})^{-1}\}. \tag{A.2}$$

Note that $\mathscr{E}(\boldsymbol{T}_{21}'\boldsymbol{T}_{21}) = \boldsymbol{I}_{p-1}$. From Lemma A.1,

$$\boldsymbol{A}_{21} \sim \frac{p-1}{n-p+2} F_{p-1,n-p+2}$$

and hence

$$\mathscr{E}(\boldsymbol{A}_{21}) = \frac{p-1}{n-p}, \quad \text{and} \quad \mathscr{E}(\boldsymbol{A}_{21})^2 = \frac{p^2-1}{(n-p)(n-p-2)}.$$

Substituting these expectations in (A.2), we obtain

$$q_p = d_p\beta_p\left(\frac{p^2-1}{(n-p)(n-p-2)} + \frac{2(p-1)}{n-p} + 1\right) + \alpha_p\,\mathrm{tr}(\boldsymbol{Q}_{11}) - d_p\alpha_p^2\mathscr{E}\,\mathrm{tr}(\boldsymbol{T}_{11}'\boldsymbol{T}_{11})^{-1}(\boldsymbol{T}_{11}'\boldsymbol{T}_{11})^{-1}. \tag{A.3}$$

Notice that $\boldsymbol{T}_{11}\boldsymbol{T}_{11}' = \boldsymbol{S}_{11}$ is distributed as $W_{p-1}(n, I)$. Therefore, from Lemma A.2,

$$\mathscr{E}\,\mathrm{tr}(\boldsymbol{T}_{11}'\boldsymbol{T}_{11})^{-1}(\boldsymbol{T}_{11}'\boldsymbol{T}_{11})^{-1} = \mathscr{E}\,\mathrm{tr}\,\boldsymbol{S}_{11}^{-2} = \frac{(n-1)(p-1)}{(n-p+1)(n-p)(n-p-2)}. \tag{A.4}$$

Substituting (A.4) in (A.3) and simplifying, we get

$$q_p = d_p\frac{n-1}{(n-p)(n-p-1)(n-p-2)} \times \left[\frac{(n-3)(n-p-1)(n-p+1)-(n-p-3)(p-1)}{(n-p+1)(n-p-1)(n-p-3)}\right] + \alpha_p\sum_{1}^{p-1} q_j. \tag{A.5}$$

From the partitioned matrix $\boldsymbol{Q}$ note that

$$q_1 = \mathscr{E}(t_{11})^{-4}[d_1 + \mathscr{E}\,\mathrm{tr}(T_{22}'T_{22})^{-1}D^*] \tag{A.6}$$

where $D^* = \mathrm{diag}(d_2, d_3, \ldots, d_p)$, and $\boldsymbol{T}_{22}\boldsymbol{T}_{22}' = \boldsymbol{S}_{22} - \boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{21}' = \boldsymbol{S}_{22.1}$, which follows a $W_{p-1}(n-1, I)$. Applying the result (A.1) to (A.6), we get

$$q_1 = \beta_1\left(d_1 + \sum_{i=2}^{p}\frac{d_i(n-2)}{(n-i-1)(n-i)}\right). \tag{A.7}$$

Since Q_{11} is similar to Q, Equation (A.5) together with (A.7) yields the inductive relations

$$q_j = \left[d_j + \left\{\sum_{l=j+1}^{p}\left(\prod_{k=j+1}^{l}\alpha_k\right)d_l\right\}\right]c_j + \alpha_j\sum_{i=1}^{j-1}q_i, \qquad j = 2, 3, \ldots, p-1, \tag{A.8}$$

where

$$c_j = \frac{n-1}{(n-j)(n-j-1)(n-j-2)} \times \left[\frac{(n-3)(n-j-1)(n-j+1)-(n-j-3)(j-1)}{(n-j+1)(n-j-1)(n-j-3)}\right].$$

ACKNOWLEDGEMENT

The authors would like to thank Alphonse Amey for programming help.

REFERENCES

Dey, D., and Srinivasan, C. (1985). Estimation of covariance matrix under Stein's loss. *Ann. Statist.*, 13, 1581–1592.

Eaton, M.L. (1970). Some problems in covariance estimation. T.R. 49, Dept. of Statistics, Stanford University.

Eaton, M.L., and Olkin, I. (1987). Best equivalent estimators of Cholesky decomposition. *Ann. Statist.*, 15, 1639–1650.

Haff, L.R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, 9, 531–544.

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

James, W., and Stein, C. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, 1, 361–379.

Kiefer, J. (1957). Invariance, minimax sequential estimation and continuous time processes. *Ann. Math. Statist.*, 28, 573–601.

Olkin, I., and Selliah, J.B. (1977). Estimating covariances in a multivariate normal distribution. *Statistical Decision Theory and Related Topics. Volume II* (S.S. Gupta, *ed.*). 313–326.

Sharma, D., and Krishnamoorthy, K. (1983). Orthogonal equivariant minimax estimators of bivariate normal covariance matrix and precision matrix. *Calcutta Statist. Assoc. Bull.*, 32, 23–45.

Sinha, B.K., and Ghosh, M. (1986). Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix, and the generalized variance under entropy loss. *Statist. Decisions*, 5, 201–227.

Smith, W.B., and Hocking, R.R. (1972). Wishart variates generator (Algorithm AS53). *Appl. Statist.*, 21, 341–345.

Tan, W.Y., and Guttman, I. (1971). A disguised Wishart variable and a related theorem. *J. Roy. Statist. Soc. Ser. B*, 33, 147–152.

Received 14 December 1987
Revised 12 January 1989
Accepted 23 January 1989

Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, Ohio 43403-0221
U.S.A.