Estimating Covariance Matrices II

by

Wei-Liem Loh
Purdue University

Technical Report #88-43

Department of Statistics
Purdue University

September 1988

AMS 1980 subject classifications: Primary 62F10, secondary 62C99

Key words and phrases: Covariance matrix, Wishart distribution, equivariant estimation, unbiased estimate of risk.

ESTIMATING COVARIANCE MATRICES II

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 $\varsigma = \Sigma_2\Sigma_1^{-1}$ under the loss functions $L_1 = \text{tr}(\hat{\varsigma} - \varsigma)'\Sigma_2^{-1}(\hat{\varsigma} - \varsigma)S_1/\text{tr}\varsigma$ and $L_2 = \text{tr}(\hat{\varsigma} - \varsigma)'\Sigma_2^{-1}(\hat{\varsigma} - \varsigma)\Sigma_1/\text{tr}\varsigma$. In this paper under the loss function L_1, we shall derive alternative estimators for ς that compare favorably with the usual estimators. We shall also show, using Monte Carlo simulation, that these estimators, suitably scaled, have excellent risk properties under L_2 with respect to the usual estimators.

1 Introduction

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 $S_2 \sim W_p(\Sigma_2, n_2)$. For simplicity, we write $\varsigma = \Sigma_2\Sigma_1^{-1}$. Estimating the eigenvalues of ς has been an area of active research in recent years. The eigenvalues of ς are important, for instance, in the problem of testing $\text{H}_0 : \Sigma_1 = \Sigma_2$ against $\text{H}_1 : \Sigma_1 \neq \Sigma_2$. The literature includes DasGupta (1986), Dey (1986), Muirhead and Verathaworn (1985) and Muirhead and Leung (1988).

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

In this paper, we consider the problem of estimating ς under the following two fully invariant loss functions:

$$\begin{aligned} L_1(\hat{\varsigma}; \varsigma, S_1, \Sigma_2) &= \text{tr}(\hat{\varsigma} - \varsigma)'\Sigma_2^{-1}(\hat{\varsigma} - \varsigma)S_1/\text{tr}\varsigma, \\ L_2(\hat{\varsigma}; \Sigma_1, \Sigma_2) &= \text{tr}(\hat{\varsigma} - \varsigma)'\Sigma_2^{-1}(\hat{\varsigma} - \varsigma)\Sigma_1/\text{tr}\varsigma. \end{aligned}$$

We observe that L_1 and L_2 are both quadratic loss functions. Also, we write

$$R_1(\hat{\varsigma}; \Sigma_1, \Sigma_2) = EL_1(\hat{\varsigma}; \varsigma, S_1, \Sigma_2),$$

$$R_2(\hat{\varsigma};\Sigma_1,\Sigma_2) = EL_2(\hat{\varsigma};\Sigma_1,\Sigma_2).$$

Here is a possible motivation for the choice of loss functions. Consider the problem of estimating the common mean of two multivariate normal populations with unknown covariance matrices. More precisely, let

$$X_i \sim N_p(\xi,\Sigma_i), \qquad S_i \sim W_p(\Sigma_i,n), \qquad i=1,2,$$

where X_1, X_2, S_1 and S_2 are mutually independent. A natural invariant loss function for estimating ξ is

$$L(\hat{\xi};\xi,\Sigma_1,\Sigma_2) = (\hat{\xi}-\xi)'(\Sigma_1^{-1}+\Sigma_2^{-1})(\hat{\xi}-\xi).$$

If Σ_1 and Σ_2 were known, the best linear unbiased estimator for ξ is

$$\hat{\xi}_0 = (\Sigma_1^{-1}+\Sigma_2^{-1})^{-1}(\Sigma_1^{-1}X_1+\Sigma_2^{-1}X_2).$$

For simplicity, we write $\Xi = (\Sigma_1^{-1}+\Sigma_2^{-1})^{-1}\Sigma_1^{-1}$. As Σ_1 and Σ_2 are unknown, following usual practice we consider estimators for ξ of the form:

$$\hat{\xi} = \hat{\Xi}X_1 + (I-\hat{\Xi})X_2,$$

where the $p\times p$ matrix $\hat{\Xi}$ is a function only of S_1 and S_2. This reduces the problem to the estimation of Ξ under the loss function

$$\begin{aligned} L(\hat{\Xi};\Sigma_1,\Sigma_2) &= \mathrm{tr}(\hat{\Xi}-\Xi)'(\Sigma_1^{-1}+\Sigma_2^{-1})(\hat{\Xi}-\Xi)(\Sigma_1+\Sigma_2) \\ &= L_2(\hat{\Xi};\Sigma_1,(\Sigma_1^{-1}+\Sigma_2^{-1})^{-1})\mathrm{tr}[(\Sigma_1^{-1}+\Sigma_2^{-1})^{-1}\Sigma_1^{-1}] \\ &\quad +\mathrm{tr}(\hat{\Xi}-\Xi)'(\Sigma_1^{-1}+\Sigma_2^{-1})(\hat{\Xi}-\Xi)\Sigma_2. \end{aligned}$$

The right hand side of the last equation is a weighted sum of two quadratic loss functions; one of which is L_2. Since L_1 is analytically more tractable than L_2, we shall use L_1 as an approximation to L_2. This, in a vague sense, is a motivation for our choice of loss functions in estimating ς.

In this paper, we shall derive alternative estimators for ς, under the loss function L_1, that compare favorably with the usual estimators. We shall also show, using Monte Carlo simulation, that these estimators, suitably scaled, have excellent risk properties under L_2 with respect to the usual estimators.

2 Equivariant Estimators

The problem we are considering is invariant under the following group of transformations:

(1) $\Sigma_i \to B\Sigma_i B'$, $\qquad S_i \to BS_iB'$ $\qquad \forall B\in GL(p,R)$, $\qquad i=1,2.$

Theorem 1 *Let $S_1 \sim W_p(\Sigma_1, n_1)$, $S_2 \sim W_p(\Sigma_2, n_2)$ with S_1, S_2 independent. Then under the group of transformations given in (1), $\hat{\varsigma}$ is an equivariant estimator for ς if and only if $\hat{\varsigma}$ can be expressed as*

$$\hat{\varsigma}(S_1, S_2) = A^{-1}\Phi(L)A,$$

where Φ is a diagonal matrix, $AS_1A' = I$, $AS_2A' = L$ and $l_1 \geq \cdots \geq l_p$ with $L = diag(l_1, \ldots l_p)$.

PROOF. Suppose $\hat{\varsigma}$ is an equivariant estimator for ς. Then

$$(2) \qquad \hat{\varsigma}(S_1, S_2) = B^{-1}\hat{\varsigma}(BS_1B', BS_2B')B \qquad \forall B \in GL(p, R).$$

We observe that $\exists A \in GL(p, R)$ such that $AS_1A' = I$ and $AS_2A' = L$, where $L = \text{diag}(l_1, \ldots, l_p)$ with $l_1 \geq \cdots \geq l_p$. Hence it follows from (2) that

$$\hat{\varsigma}(S_1, S_2) = A^{-1}\hat{\varsigma}(I, L)A.$$

By invariance again, we have

$$\hat{\varsigma}(I, L) = D^{-1}\hat{\varsigma}(I, L)D, \qquad \forall D = \text{diag}(\pm 1).$$

This implies that $\hat{\varsigma}$ is diagonal. Writing

$$\Phi(L) = \hat{\varsigma}(I, L),$$

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

3 Calculus on Eigenstructure

Let $S_1 \sim W_p(\Sigma_1, n_1)$ and $S_2 \sim W_p(\Sigma_2, n_2)$. For simplicity we write:

$$s_{jk}^{(i)} = (S_i)_{jk}, \qquad \tilde{\nabla}_{jk}^{(i)} = (1/2)(1 + \delta_{jk})\partial/\partial s_{jk}^{(i)} \qquad \forall i, j, k,$$

where δ_{jk} denotes the Kronecker delta. We observe that $\exists A \in GL(p, R)$ such that $AS_1A' = I$ and $AS_2A' = L$ where $L = \text{diag}(l_1, \ldots l_p)$ with $l_1 \geq \cdots \geq l_p$. In this section, the partial derivatives of A^{-1} and L with respect to S_1 and S_2 are computed.

Proposition 1 *Let $S_1 \sim W_p(\Sigma_1, n_1)$ and $S_2 \sim W_p(\Sigma_2, n_2)$. Then with L, $A = (a_{il})$, $A^{-1} = (a^{il})$ as defined above, we have*

$$\begin{aligned}
\tilde{\nabla}^{(1)}_{jk} l_i &= -l_i a_{ij} a_{ik}, \\
\tilde{\nabla}^{(2)}_{jk} l_i &= a_{ij} a_{ik}, \\
\tilde{\nabla}^{(1)}_{jk} a^{il} &= \frac{1}{2} a^{il} a_{lj} a_{lk} + \frac{1}{2} \sum_{i' \neq l} a^{ii'} (a_{i'j} a_{lk} + a_{i'k} a_{lj}) \frac{l_{i'}}{l_{i'} - l_l}, \\
\tilde{\nabla}^{(2)}_{jk} a^{il} &= \frac{1}{2} \sum_{i' \neq l} a^{ii'} (a_{i'j} a_{lk} + a_{i'k} a_{lj}) \frac{1}{l_l - l_{i'}}.
\end{aligned}$$

PROOF. On differentiating $S_1 = A^{-1} A'^{-1}$ and $S_2 = A^{-1} L A'^{-1}$, we have

$$\begin{aligned}
\mathrm{d}S_1 &= A^{-1}(\mathrm{d}A'^{-1}) + (\mathrm{d}A^{-1})A'^{-1}, \\
\mathrm{d}S_2 &= A^{-1}L(\mathrm{d}A'^{-1}) + (\mathrm{d}A^{-1})LA'^{-1} + A^{-1}(\mathrm{d}L)A'^{-1}.
\end{aligned}$$

Multiplying these equations by A on the left and A' on the right we get

$$A(\mathrm{d}S_1)A' = (\mathrm{d}A'^{-1})A' + A(\mathrm{d}A^{-1}), \tag{3}$$

$$A(\mathrm{d}S_2)A' = L(\mathrm{d}A'^{-1})A' + A(\mathrm{d}A^{-1})L + (\mathrm{d}L). \tag{4}$$

CASE I. Suppose $\mathrm{d}S_2 = 0$. From (3) and (4), we arrive at

$$\mathrm{d}L = LA(\mathrm{d}A^{-1}) - A(\mathrm{d}A^{-1})L - LA(\mathrm{d}S_1)A'; \tag{5}$$

which implies that

$$\begin{aligned}
\mathrm{d}l_i &= -l_i[A(\mathrm{d}S_1)A']_{ii} \\
&= -\sum_{j,k} l_i a_{ij} (\mathrm{d}s^{(1)}_{jk}) a_{ik}.
\end{aligned}$$

Thus we conclude that

$$\tilde{\nabla}^{(1)}_{jk} l_i = -l_i a_{ij} a_{ik}.$$

For $i \neq j$, it follows from (5) that

$$\begin{aligned}
(\mathrm{d}L)_{ij} &= 0 \\
&= l_i[A(\mathrm{d}A^{-1})]_{ij} - [A(\mathrm{d}A^{-1})]_{ij} l_j - l_i[A(\mathrm{d}S_1)A']_{ij}.
\end{aligned}$$

This reduces to

$$[A(\mathrm{d}A^{-1})]_{ij} = \frac{l_i}{l_i - l_j} \sum_{k,l} a_{ik} (\mathrm{d}s^{(1)}_{kl}) a_{jl}. \tag{6}$$

Also we observe from (3) that

$$[A(\mathrm{d}A^{-1})]_{ii} = \frac{1}{2}[A(\mathrm{d}S_1)A']_{ii}. \tag{7}$$

Now it follows from (6) and (7) that

$$\begin{aligned}(\mathrm{d}A^{-1})_{ij} &= [A^{-1}A(\mathrm{d}A^{-1})]_{ij} \\ &= \sum_{i'} a^{ii'}[A(\mathrm{d}A^{-1})]_{i'j} \\ &= \frac{1}{2}a^{ij}\sum_{k,l} a_{jk}(\mathrm{d}s_{kl}^{(1)})a_{jl} + \frac{1}{2}\sum_{i'\neq j} a^{ii'}\frac{l_{i'}}{l_{i'}-l_j}\sum_{k,l} a_{i'k}(\mathrm{d}s_{kl}^{(1)})a_{jl}.\end{aligned}$$

Hence we conclude that

$$\tilde{\nabla}_{jk}^{(1)}a^{il} = \frac{1}{2}a^{il}a_{lj}a_{lk} + \frac{1}{2}\sum_{i'\neq l} a^{ii'}(a_{i'j}a_{lk} + a_{i'k}a_{lj})\frac{l_{i'}}{l_{i'}-l_l}.$$

Case II. Suppose $\mathrm{d}S_2 = 0$. The proof of

$$\begin{aligned}\tilde{\nabla}_{jk}^{(2)}l_i &= a_{ij}a_{ik}, \\ \tilde{\nabla}_{jk}^{(2)}a^{il} &= \frac{1}{2}\sum_{i'\neq l} a^{ii'}(a_{i'j}a_{lk} + a_{i'k}a_{lj})\frac{1}{l_l - l_{i'}}\end{aligned}$$

is similar to that of Case I and is omitted. □

4 Two Identities

We shall now state two identities which are needed in the sequel. Their proofs are given in Loh (1988) and hence are omitted.

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, then g is almost differentiable if each of its coordinate functions are.

Theorem 2 (Normal Identity) *Let $X = (X_1, \ldots, X_p)' \sim N_p(\xi, \Sigma)$ and $g : R^p \to R^p$ be an almost differentiable function such that $E[\sum_{i,j} \mid \partial g_i(X) / \partial X_j \mid]$ is finite. Then*

$$E[\Sigma^{-1}(X-\xi)g'(X)] = E[\nabla g'(X)],$$

where $\nabla = (\partial/\partial X_1, \ldots, \partial/\partial X_p)'$.

The Normal identity was first proved by Stein (1973).

Let $\mathcal{S}_p$ denote the set of $p \times p$ positive definite matrices. Also we write for $1 \le i, j \le p$,

$$\tilde{\nabla} = (\tilde{\nabla}_{ij})_{p \times p}, \text{ where } \tilde{\nabla}_{ij} = (1/2)(1+\delta_{ij})\partial/\partial s_{ij},$$

where δ_{ij} denotes the Kronecker delta.

Theorem 3 (Wishart Identity) *Let $X = (X_1, \ldots, X_n)$ 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. We suppose $n \ge p$. Let $g : \mathcal{S}_p \to R^{p \times p}$ be such that $x \mapsto g(xx') : R^{p \times n} \to R^{p \times p}$ is almost differentiable. Then, with $S = XX'$, we have*

$$Etr\Sigma^{-1} g(S) = Etr[(n-p-1)S^{-1}g(S) + 2\tilde{\nabla} g(S)],$$

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

The Wishart identity was proved by Stein (1975) and Haff (1977) independently.

5 Unbiased Estimate of Risk

In this section, we shall compute, up to a constant which depends only on the parameters, the unbiased estimate of the risk of an almost arbitrary equivariant estimator of ς $(= \Sigma_2 \Sigma_1^{-1})$. First, we start with a lemma.

Lemma 1 *Let A, $\Phi = diag(\phi_1, \ldots, \phi_p)$ and $L = diag(l_1, \ldots, l_p)$ be defined as in Theorem 1. Then*

$$\begin{aligned} tr\tilde{\nabla}^{(1)}(A^{-1}\Phi A'^{-1}) &= \sum_i (\phi_i - l_i \frac{\partial \phi_i}{\partial l_i} + \phi_i \sum_{j \ne i} \frac{l_j}{l_j - l_i}), \\ tr\tilde{\nabla}^{(2)}(A^{-1}\Phi^2 A'^{-1}) &= \sum_i (2\phi_i \frac{\partial \phi_i}{\partial l_i} + \phi_i^2 \sum_{j \ne i} \frac{1}{l_i - l_j}). \end{aligned}$$

PROOF.

$$\begin{aligned} & \mathrm{tr}\tilde{\nabla}^{(1)}(A^{-1}\Phi A'^{-1}) \\ = & \sum_{i,j,k} \tilde{\nabla}_{ij}^{(1)}(a^{jk}\phi_k a^{ik}) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i,j,k}[(\tilde{\nabla}_{ij}^{(1)}a^{jk})\phi_k a^{ik} + a^{jk}(\tilde{\nabla}_{ij}^{(1)}\phi_k)a^{ik} + a^{jk}\phi_k(\tilde{\nabla}_{ij}^{(1)}a^{ik})] \\
&= \sum_{i,j,k}[2\phi_k a^{ik}(\tilde{\nabla}_{ij}^{(1)}a^{jk}) + a^{jk}a^{ik}\sum_m(\tilde{\nabla}_{ij}^{(1)}l_m)\frac{\partial\phi_k}{\partial l_m}].
\end{aligned}
$$

Now it follows from Proposition 1 that

$$
\begin{aligned}
&\mathrm{tr}\tilde{\nabla}^{(1)}(A^{-1}\Phi A'^{-1}) \\
&= \sum_{i,j,k}\{\phi_k a^{ik}[a^{jk}a_{ki}a_{kj} + \sum_{i'\neq k}a^{ji'}(a_{i'i}a_{kj} + a_{i'j}a_{ki})\frac{l_{i'}}{l_{i'}-l_k}] \\
&\quad -a^{jk}a^{ik}\sum_m a_{mi}a_{mj}l_m\frac{\partial\phi_k}{\partial l_m}\} \\
&= \sum_i(\phi_i - l_i\frac{\partial\phi_i}{\partial l_i} + \phi_i\sum_{j\neq i}\frac{l_j}{l_j-l_i}).
\end{aligned}
$$

The second part of this lemma can be proved similarly. □

With this lemma, we shall now prove the main result of this section.

Theorem 4 *Let $\hat{\varsigma}$ be an estimator for ς where*

$$\hat{\varsigma}(S_1, S_2) = A^{-1}\Phi(L)A,$$

$\Phi = diag(\phi_1,\ldots,\phi_p)$, $AS_1A' = I$, $AS_2A' = L = diag(l_1,\ldots,l_p)$ with $l_1 \geq \cdots \geq l_p$. Suppose Φ satisfies the conditions of the Wishart identity in the sense that

$$
\begin{aligned}
E\mathrm{tr}(\Sigma_1^{-1}A^{-1}\Phi A'^{-1}) &= E\mathrm{tr}[2\tilde{\nabla}^{(1)}(A^{-1}\Phi A'^{-1}) + (n_1-p-1)\Phi], \\
E\mathrm{tr}(\Sigma_2^{-1}A^{-1}\Phi^2 A'^{-1}) &= E\mathrm{tr}[2\tilde{\nabla}^{(2)}(A^{-1}\Phi^2 A'^{-1}) + (n_2-p-1)L^{-1}\Phi^2].
\end{aligned}
$$

Then under the loss function L_1, the risk of $\hat{\varsigma}$ is given by

$$
\begin{aligned}
R_1(\hat{\varsigma};\Sigma_1,\Sigma_2) &= n_1 + E\sum_i[\frac{n_2-p-1}{l_i}\phi_i^2 + 2\phi_i^2\sum_{j\neq i}\frac{1}{l_i-l_j} + 4\phi_i\frac{\partial\phi_i}{\partial l_i} \\
&\quad -2(n_1-p+1)\phi_i + 4\phi_i\sum_{j\neq i}\frac{l_j}{l_i-l_j} + 4l_i\frac{\partial\phi_i}{\partial l_i}]/\mathrm{tr}\varsigma.
\end{aligned}
$$

PROOF. We observe from the Wishart identity that

$$
\begin{aligned}
E\mathrm{tr}(\Sigma_1^{-1}A^{-1}\Phi A'^{-1}) &= E\mathrm{tr}[2\tilde{\nabla}^{(1)}(A^{-1}\Phi A'^{-1}) + (n_1-p-1)\Phi], \\
E\mathrm{tr}(\Sigma_2^{-1}A^{-1}\Phi^2 A'^{-1}) &= E\mathrm{tr}[2\tilde{\nabla}^{(2)}(A^{-1}\Phi^2 A'^{-1}) + (n_2-p-1)L^{-1}\Phi^2].
\end{aligned}
$$

Now it follows from Lemma 1 that

$$
\begin{aligned}
E\mathrm{tr}(\Sigma_1^{-1}A^{-1}\Phi A'^{-1}) &= E\sum_i[(n_1-p-1)\phi_i - 2\phi_i\sum_{j\neq i}\frac{l_j}{l_i-l_j} \\
&\quad +2\phi_i - 2l_i\frac{\partial\phi_i}{\partial l_i}], \qquad (8)\\
E\mathrm{tr}(\Sigma_2^{-1}A^{-1}\Phi^2 A'^{-1}) &= E\sum_i[\frac{n_2-p-1}{l_i}\phi_i^2 + 2\phi_i^2\sum_{j\neq i}\frac{1}{l_i-l_j} \\
&\quad +4\phi_i\frac{\partial\phi_i}{\partial l_i}]. \qquad (9)
\end{aligned}
$$

Finally the risk of $\hat{\varsigma}$ is given by

$$
\begin{aligned}
R_1(\hat{\varsigma};\Sigma_1,\Sigma_2) &= EL_1(\hat{\varsigma};\varsigma,S_1,\Sigma_2)\\
&= E\mathrm{tr}(\hat{\varsigma}-\varsigma)'\Sigma_2^{-1}(\hat{\varsigma}-\varsigma)S_1/\mathrm{tr}\varsigma\\
&= E\mathrm{tr}(\hat{\varsigma}'\Sigma_2^{-1}\hat{\varsigma}S_1+\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}S_1-2\Sigma_1^{-1}\hat{\varsigma}S_1)/\mathrm{tr}\varsigma\\
&= E\mathrm{tr}(\Sigma_2^{-1}A^{-1}\Phi^2A'^{-1}+n_1\Sigma_2\Sigma_1^{-1}\\
&\quad -2\Sigma_1^{-1}A^{-1}\Phi A'^{-1})/\mathrm{tr}\varsigma.
\end{aligned}
$$

It follows from (8) and (9) that

$$
\begin{aligned}
R_1(\hat{\varsigma};\Sigma_1,\Sigma_2) &= n_1 + E\sum_i[\frac{n_2-p-1}{l_i}\phi_i^2 + 2\phi_i^2\sum_{j\neq i}\frac{1}{l_i-l_j} + 4\phi_i\frac{\partial\phi_i}{\partial l_i}\\
&\quad -2(n_1-p+1)\phi_i + 4\phi_i\sum_{j\neq i}\frac{l_j}{l_i-l_j} + 4l_i\frac{\partial\phi_i}{\partial l_i}]/\mathrm{tr}\varsigma.
\end{aligned}
$$

This completes the proof. □

6 Usual Estimators

The usual estimators for ς are of the form $cS_2S_1^{-1}$ where c is a constant. This class of estimators includes the maximum likelihood estimator $\hat{\varsigma}^{ML} = (n_1/n_2)S_2S_1^{-1}$ and also the uniformly minimum variance unbiased estimator $\hat{\varsigma}^{UB} = [(n_1-p-1)/n_2]S_2S_1^{-1}$. We define the best usual estimator to be that usual estimator which minimizes the risk among the usual estimators.

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, S_2 independent. With respect to the loss function L_1, the best usual estimator $\hat{\varsigma}^{BU}$ for ς is $[(n_1-p-1)/(n_2+p+1)]S_2S_1^{-1}$. Furthermore, the risk of $\hat{\varsigma}^{BU}$ is*

$$R_1(\hat{\varsigma}^{BU};\Sigma_1,\Sigma_2) = (p+1)(n_1+n_2)/(n_2+p+1).$$

PROOF. Under L_1, the risk of estimators of the form $cS_2S_1^{-1}$ is given by

$$\begin{aligned} & R_1(cS_2S_1^{-1};\Sigma_1,\Sigma_2) \\ = \; & E\mathrm{tr}(cS_2S_1^{-1}-\Sigma_2\Sigma_1^{-1})'\Sigma_2^{-1}(cS_2S_1^{-1}-\Sigma_2\Sigma_1^{-1})S_1/\mathrm{tr}\varsigma \\ = \; & \frac{n_2(n_2+p+1)}{n_1-p-1}c^2-2n_2c+n_1. \end{aligned}$$

This is minimized when $c=(n_1-p-1)/(n_2+p+1)$. □

REMARK. We observe from the proof of Theorem 5 that for all Σ_1 and Σ_2,

$$R_1(\hat{\varsigma}^{BU};\Sigma_1,\Sigma_2)\le R_1(\hat{\varsigma}^{UB};\Sigma_1,\Sigma_2)\le R_1(\hat{\varsigma}^{ML};\Sigma_1,\Sigma_2).$$

Theorem 6 *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. With respect to the loss function L_2, the best usual estimator for ς is*

$$\hat{\varsigma}^{BU}=\frac{(n_1-p)(n_1-p-3)}{(n_1-1)(n_2+p+1)}S_2S_1^{-1}.$$

Furthermore, the risk of $\hat{\varsigma}^{BU}$ is

$$R_2(\hat{\varsigma}^{BU};\Sigma_1,\Sigma_2)=1-\frac{(n_1-p)(n_1-p-3)n_2}{(n_1-1)(n_1-p-1)(n_2+p+1)}.$$

PROOF. Under L_2, the risk of estimators of the form $cS_2S_1^{-1}$ is given by

$$\begin{aligned} & R_2(cS_2S_1^{-1};\Sigma_1,\Sigma_2) \\ = \; & E\mathrm{tr}(cS_2S_1^{-1}-\Sigma_2\Sigma_1^{-1})'\Sigma_2^{-1}(cS_2S_1^{-1}-\Sigma_2\Sigma_1^{-1})\Sigma_1/\mathrm{tr}\varsigma \\ = \; & \frac{(n_1-1)(n_2+p+1)n_2}{(n_1-p)(n_1-p-1)(n_1-p-3)}c^2-\frac{2n_2}{n_1-p-1}c+1. \end{aligned}$$

This is minimized by $c=(n_1-p)(n_1-p-3)/[(n_1-1)((n_2+p+1)]$. □

REMARK. We observe from the proof of Theorem 6 that for all Σ_1 and Σ_2,

$$R_2(\hat{\varsigma}^{BU};\Sigma_1,\Sigma_2)\le R_2(\hat{\varsigma}^{UB};\Sigma_1,\Sigma_2)\le R_2(\hat{\varsigma}^{ML};\Sigma_1,\Sigma_2).$$

7 Alternative Estimators

It is well-known that the eigenvalues of $S_2S_1^{-1}$ are more spread out than the eigenvalues of its expectation. By correcting for this eigenvalue distortion, we derive alternative estimators for ς which compare favorably with the best usual estimator $\hat{\varsigma}^{BU}$ under L_1 loss. Furthermore, these estimators give substantial savings in risk when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are close together.

7.1 Adjusted Usual Estimator

Under L_1 loss, the best usual estimator for ς can be written as

$$\begin{aligned}\hat{\varsigma}^{BU} &= [(n_1-p-1)/(n_2+p+1)]S_2S_1^{-1} \\ &= A^{-1}\Phi^{BU}A,\end{aligned}$$

where the j'th diagonal element of the diagonal matrix Φ^{BU} is $l_j(n_1-p-1)/(n_2+p+1)$. A natural way to improve on this estimator would be to consider estimators of the form $\hat{\varsigma} = A^{-1}\Phi A$, where for some constants c_j, $j = 1,\ldots,p$, the j'th diagonal element of the diagonal matrix Φ is $\phi_j = c_jl_j$. We define the adjusted usual estimator to be

$$\hat{\varsigma}^{AU} = A^{-1}\Phi^{AU}A$$

where, for $j = 1,\ldots,p$, the j'th diagonal element of the diagonal matrix Φ^{AU} is $\phi_j^{AU} = l_j(n_1-p-1)/(n_2+p+3-2j)$. We shall show in this subsection that $\hat{\varsigma}^{AU}$ dominates $\hat{\varsigma}^{BU}$ with respect to the loss function L_1. First we need a lemma.

Lemma 2 *For each i, $1 \le i \le p$, we have*

$$\sum_{j<i}\left(\frac{1}{n_2+p+3-2j}\right)^2 \le \frac{i-1}{(n_2+p+1)(n_2+p+3-2i)}.$$

PROOF. This follows directly from mathematical induction on i. □

Theorem 7 *With respect to the loss function L_1, $\hat{\varsigma}^{AU}$ dominates $\hat{\varsigma}^{BU}$.*

PROOF. For simplicity, we write $c_j = (n_1-p-1)/(n_2+p+3-2j)$ for $1 \le j \le p$. We observe that Φ^{AU} satisfies the conditions of the Wishart identity in the sense of Theorem 4. Hence it follows from Theorem 4 that

$$\begin{aligned}&R_1(\hat{\varsigma}^{AU};\Sigma_1,\Sigma_2) \\ =\;& n_1 + E\sum_i[(n_2-p+3)c_i^2l_i + 2\sum_{j<i}\frac{c_j^2(l_i^2-l_j^2)}{l_i-l_j} + 2\sum_{j<i}\frac{l_i^2(c_i^2-c_j^2)}{l_i-l_j} \\ & -2(n_1-p-1)c_il_i + 4\sum_{j>i}\frac{l_il_j(c_i-c_j)}{l_i-l_j}]/\mathrm{tr}\varsigma \\ \le\;& n_1 + E\sum_i[(n_2-p+3)c_i^2l_i - 2(n_1-p-1)c_il_i + 2\sum_{j<i}c_j^2(l_i+l_j)]/\mathrm{tr}\varsigma\end{aligned}$$

$$
\begin{aligned}
&= n_1 + E\sum_i[(n_2+p+3-2i)c_i^2 - 2(n_1-p-1)c_i + 2\sum_{j<i} c_j^2]l_i/\mathrm{tr}\varsigma \\
&= R_1(\hat{\varsigma}^{BU};\Sigma_1,\Sigma_2) - 2(n_1-p-1)^2 E\sum_i[\frac{i-1}{(n_2+p+1)(n_2+p+3-2i)} \\
&\quad -\sum_{j<i}(\frac{1}{n_2+p+3-2j})^2]l_i/\mathrm{tr}\varsigma \\
&\le R_1(\hat{\varsigma}^{BU};\Sigma_1,\Sigma_2).
\end{aligned}
$$

The second last equality follows from Theorem 5 and the last inequality follows from Lemma 2. This completes the proof. □

7.2 Berger-type Estimators

In this subsection, we shall use a technique of Berger (1980) to derive a class of estimators for ς which dominates $\hat{\varsigma}^{AU}$ under L_1 loss. First we need some additional notation. We let

$$\hat{\varsigma}^{BE} = A^{-1}\Phi^{BE}A,$$

where $\Phi^{BE} = \mathrm{diag}(\phi_1^{BE},\ldots,\phi_p^{BE})$ with

$$
\begin{aligned}
\phi_i^{BE} &= \frac{n_1-p-1}{n_2+p+3-2i}l_i + \frac{c}{b+u}, \\
u &= \sum_j[\frac{n_2+p+3-2j}{(n_1-p-1)l_j}]^2,
\end{aligned}
$$

$c : R^+ \to R$ being a suitable function of u and b being a suitable positive constant. First we state two rather technical lemmas.

Lemma 3

$$\max_{\vec{y}} \frac{\sum_i y_i}{b+\sum_i y_i^2} = \frac{\sqrt{p}}{2\sqrt{b}}.$$

PROOF. This follows easily by taking partial derivatives with respect to the y_i's. □

Lemma 4 *Let $d_j = (n_1-p-1)/(n_2+p+3-2j)$ for $1 \le j \le p$. Then if $p \ge 3$, $n_1 \ge p$ and $n_2 \ge p$, we have*

$$(\sum_{i=1}^{p-1} d_i) - d_p \ge 0.$$

PROOF. We observe that

$$(10)\ d_p - d_{p-1} - d_{p-2} = \frac{(n_1-p-1)[-n_2^2+2n_2(p-3)-p^2+6p-1]}{(n_2-p+3)(n_2-p+5)(n_2-p+7)}.$$

Maximum of $-n_2^2+2n_2(p-3)-p^2+6p-1$ occurs at $n_2 = p-3$. Since $n_1 \geq p$ and $n_2 \geq p$, by taking $n_2 = p$ in the right hand side of (10) we conclude that

$$\begin{aligned} d_p - d_{p-1} - d_{p-2} &\leq -\frac{n_1-p-1}{(n_2-p+3)(n_2-p+5)(n_2-p+7)} \\ &\leq 0. \end{aligned}$$

This completes the proof. □

Now we shall prove the main result of this subsection.

Theorem 8 *With the above notation, $\hat{\varsigma}^{BE}$ dominates $\hat{\varsigma}^{AU}$ in estimating ς under L_1 loss whenever*

1. $p \geq 3,\ n_1 \geq p,\ n_2 \geq p,$
2. $c(u) \geq 0,\ c'(u) \geq 0$ *for all* $u \geq 0,$
3. $\sup_u c(u)/\sqrt{b} \leq 4(p^2+p-4)(n_2-p+3)/[\sqrt{p}(n_1-p-1)(n_2-p+7)].$

PROOF. For simplicity, we write for $1 \leq j \leq p$,

$$\begin{aligned} d_j &= (n_1-p-1)/(n_2+p+3-2j), \\ \alpha_j &= c/[d_j l_j (b+u)]. \end{aligned}$$

Then taking partial derivatives, we have

$$\frac{\partial \alpha_j}{\partial l_j} = \frac{1}{d_j l_j^2 (b+u)}\left[\frac{2c}{d_j^2 l_j^2 (b+u)} - c - \frac{2c'}{d_j^2 l_j^2}\right].$$

Next we observe that

$$(11)\qquad \sum_{j>i} \frac{d_i^2\alpha_i^2 l_i^2 - d_j^2\alpha_j^2 l_j^2}{l_i - l_j} = 0.$$

Also,

$$\sum_i \left(\sum_{j>i} \frac{d_i^2\alpha_i l_i^2 - d_j^2\alpha_j l_j^2}{l_i - l_j} + \sum_{j>i} \frac{d_i\alpha_i l_i l_j - d_j\alpha_j l_i l_j}{l_i - l_j}\right)$$

$$= \frac{c}{b+u}\sum_i[\sum_{j>i}\frac{d_il_i - d_jl_j}{l_i - l_j} - (p-i)]$$

$$= \frac{c}{b+u}\{\sum_i\sum_{j>i}[\frac{d_i(l_i-l_j)}{l_i-l_j} + \frac{l_j(d_i-d_j)}{l_i-l_j}] - \frac{p^2-p}{2}\}$$

$$(12) \qquad = \frac{c}{b+u}[\frac{p-p^2}{2} + \sum_i d_i(p-i) + \sum_i\sum_{j>i}\frac{l_j(d_i-d_j)}{l_i-l_j}].$$

Furthermore we observe that

$$\sum_i(d_i^2+d_i)l_i^2(\partial\alpha_i/\partial l_i)$$

$$\leq \frac{c}{b+u}\sum_i[-1+\frac{2}{d_i^2l_i^2(b+u)} - d_i + \frac{2d_i}{d_i^2l_i^2(b+u)}]$$

$$\leq \frac{c}{b+u}[-(p-2) - \sum_i d_i + 2d_p\sum_i\frac{1}{d_i^2l_i^2(b+u)}]$$

$$\leq -\frac{c}{b+u}[p-2-d_p+\sum_{i=1}^{p-1}d_i]$$

$$(13) \qquad \leq -\frac{c}{b+u}(p-2).$$

The first inequality follows from the assumption $c'(u) \geq 0$, the second inequality follows from $d_1 < \cdots < d_p$ and the last inequality follows from Lemma 4. Finally, since $c'(u) \geq 0$, we have

$$\sum_i[(n_2-p+3)d_i^2l_i\alpha_i^2 + 4d_i^2l_i^2\alpha_i(\partial\alpha_i/\partial l_i)]$$

$$\leq \frac{c^2}{(b+u)^2}[\frac{8}{l_p} + (n_2-p-1)\sum_i\frac{1}{l_i}]$$

$$\leq (n_2-p+7)\frac{c^2d_p\sum_i(d_il_i)^{-1}}{(b+u)^2}$$

$$\leq \frac{(n_1-p-1)(n_2-p+7)c^2}{(n_2-p+3)(b+u)}\max_{\vec{y}}\frac{\sum_i y_i}{b+\sum_i y_i^2}$$

$$(14) \qquad \leq \frac{(n_1-p-1)(n_2-p+7)c^2\sqrt{p}}{2(n_2-p+3)(b+u)\sqrt{b}}.$$

The last inequality follows from Lemma 3. We observe that Φ^{BE} satisfies the conditions of the Wishart identity in the sense of Theorem 4. Hence from Theorem 4 we have

$$
\begin{aligned}
& R_1(\hat{\varsigma}^{BE};\Sigma_1,\Sigma_2) - R_1(\hat{\varsigma}^{AU};\Sigma_1,\Sigma_2) \\
= \; & E\sum_i[\frac{(n_1-p-1)^2(n_2-p-1)}{(n_2+p+3-2i)^2}l_i(2\alpha_i+\alpha_i^2) \\
& +4(\frac{n_1-p-1}{n_2+p+3-2i})^2 l_i(l_i\frac{\partial\alpha_i}{\partial l_i}+l_i\alpha_i\frac{\partial\alpha_i}{\partial l_i}+2\alpha_i+\alpha_i^2) \\
& +2(\frac{n_1-p-1}{n_2+p+3-2i})^2(2\alpha_i+\alpha_i^2)\sum_{j\neq i}\frac{l_i^2}{l_i-l_j} \\
& -2\frac{(n_1-p+1)(n_1-p-1)}{n_2+p+3-2i}l_i\alpha_i \\
& +4(\frac{n_1-p-1}{n_2+p+3-2i})\alpha_i\sum_{j\neq i}\frac{l_il_j}{l_i-l_j}+4(\frac{n_1-p-1}{n_2+p+3-2i})l_i(\alpha_i+l_i\frac{\partial\alpha_i}{\partial l_i}) \\
= \; & E\sum_i[(n_2-p+3)d_i^2l_i\alpha_i^2+4(d_i^2+d_i)l_i^2\frac{\partial\alpha_i}{\partial l_i}+4d_i^2l_i^2\alpha_i\frac{\partial\alpha_i}{\partial l_i} \\
& -4l_i\alpha_id_i^2(p-i)+2\sum_{j>i}\frac{d_i^2\alpha_i^2l_i^2-d_j^2\alpha_j^2l_j^2}{l_i-l_j} \\
& +4\sum_{j>i}\frac{d_i^2\alpha_il_i^2-d_j^2\alpha_jl_j^2}{l_i-l_j}+4\sum_{j>i}\frac{d_i\alpha_il_il_j-d_j\alpha_jl_il_j}{l_i-l_j}].
\end{aligned}
$$

It follows from (11) to (14) that

$$
\begin{aligned}
& R_1(\hat{\varsigma}^{BE};\Sigma_1,\Sigma_2) - R_1(\hat{\varsigma}^{AU};\Sigma_1,\Sigma_2) \\
\leq \; & E\frac{c}{b+u}[\frac{(n_1-p-1)(n_2-p+7)c\sqrt{p}}{2(n_2-p+3)\sqrt{b}}-2(p^2-p) \\
& -4(p-2)+4\sum_i\sum_{j>i}\frac{l_j(d_i-d_j)}{l_i-l_j}] \\
\leq \; & E\frac{c}{b+u}[\frac{(n_1-p-1)(n_2-p+7)c\sqrt{p}}{2(n_2-p+3)\sqrt{b}}-2(p^2+p-4)] \\
\leq \; & 0.
\end{aligned}
$$

The last inequality follows from assumption 3. This completes the proof. □

7.3 Stein-type Estimators

By an approximate minimization of the unbiased estimate of the risk of an almost arbitrary orthogonally invariant estimator of a covariance matrix, Stein (1975) derived an estimator whose risk compares very favorably with the minimax risk under Stein's loss. 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 alternative equivariant estimators $\hat{\varsigma}^{S1}$ and $\hat{\varsigma}^{S2}$ for ς under the loss function L_1. The construction is as follows:

Let $\hat{\varsigma}$ be an estimator for ς where

$$\hat{\varsigma}(S_1, S_2) = A^{-1}\Phi(L)A,$$

$\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_p)$, $AS_1A' = I$, $AS_2A' = L = \mathrm{diag}(l_1, \ldots, l_p)$ with $l_1 \geq \cdots \geq l_p$. Under loss function L_1, we observe from Theorem 4 that

$$\begin{aligned} R_1(\hat{\varsigma}; \Sigma_1, \Sigma_2) &= n_1 + E\sum_i [\frac{n_2 - p - 1}{l_i}\phi_i^2 + 2\phi_i^2 \sum_{j \neq i} \frac{1}{l_i - l_j} + 4\phi_i \frac{\partial \phi_i}{\partial l_i} \\ &\quad -2(n_1 - p + 1)\phi_i + 4\phi_i \sum_{j \neq i} \frac{l_j}{l_i - l_j} + 4l_i \frac{\partial \phi_i}{\partial l_i}]/\mathrm{tr}\varsigma. \end{aligned}$$

This is equivalent to

$$\begin{aligned} &[R_1(\hat{\varsigma}; \Sigma_1, \Sigma_2) - n_1]\mathrm{tr}\varsigma \\ = \;& E\sum_i [\frac{n_2 - p + 3}{l_i}\phi_i^2 + 2\phi_i^2 \sum_{j \neq i} \frac{1}{l_i - l_j} + 4l_i\phi_i \frac{\partial}{\partial l_i}(\frac{\phi}{l_i}) \\ & -2(n_1 - p - 1)\phi_i + 4\phi_i \sum_{j \neq i} \frac{l_j}{l_i - l_j} + 4l_i^2 \frac{\partial}{\partial l_i}(\frac{\phi}{l_i})] \\ = \;& E\tilde{R}, \end{aligned}$$

where

$$\begin{aligned} \tilde{R} &= \sum_i [\frac{n_2 - p + 3}{l_i}\phi_i^2 + 2\phi_i^2 \sum_{j \neq i} \frac{1}{l_i - l_j} + 4l_i\phi_i \frac{\partial}{\partial l_i}(\frac{\phi}{l_i}) \\ &\quad -2(n_1 - p - 1)\phi_i + 4\phi_i \sum_{j \neq i} \frac{l_j}{l_i - l_j} + 4l_i^2 \frac{\partial}{\partial l_i}(\frac{\phi}{l_i})]. \end{aligned}$$

By ignoring the derivative terms in $\tilde{R}$, we get

$$\begin{aligned} \hat{R} &= \sum_i [\frac{n_2 - p + 3}{l_i}\phi_i^2 + 2\phi_i^2 \sum_{j \neq i} \frac{1}{l_i - l_j} \\ &\quad -2(n_1 - p - 1)\phi_i + 4\phi_i \sum_{j \neq i} \frac{l_j}{l_i - l_j}]. \end{aligned}$$

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

$$\begin{aligned}\phi_i &= l_i(n_1 - p - 1 - 2\sum_{j\neq i}\frac{l_j}{l_i - l_j})/(n_2 + p + 1 + 2\sum_{j\neq i}\frac{l_j}{l_i - l_j}) \\ &= (l_i/\alpha_i)(1/\beta_i)^{-1}, \text{ say,} \qquad (15)\end{aligned}$$

where for $1 \leq i \leq p$,

$$\begin{aligned}\alpha_i &= n_2 + p + 1 + 2\sum_{j\neq i}\frac{l_j}{l_i - l_j}, \\ \beta_i &= n_1 - p - 1 - 2\sum_{j\neq i}\frac{l_j}{l_i - l_j}.\end{aligned}$$

We note that the ϕ_i's should follow a natural ordering. That is

$$\phi_1 \geq \cdots \geq \phi_p \geq 0.$$

However with the ϕ_i's defined in (15), this natural ordering may be altered. Here are two methods for correcting this. In both methods, the use of Stein's (1975) isotonic regression is crucial. For a detailed description of Stein's isotonic regression, see for example Lin and Perlman (1985).

METHOD I.

Let $\varphi_i = l_i/\alpha_i$ for $1 \leq i \leq p$. As it is, the condition $\varphi_1 \geq \cdots \geq \varphi_p \geq 0$ may not be satisfied. By applying Stein's isotonic regression to the φ_i's, we arrive at a new set of φ_i's, denoted by φ_i^{ST}, $i = 1, \ldots, p$ such that $\varphi_1^{ST} \geq \cdots \geq \varphi_p^{ST} \geq 0$. Writing the adjusted usual estimator as $\hat{\varsigma}^{AU} = A^{-1}\Phi^{AU}A$ where

$$\phi_i^{AU} = [l_i/(n_2 + p + 3 - 2i)][1/(n_1 - p - 1)]^{-1}$$

and comparing it with (15), suggest replacing the factor $(1/\beta_i)^{-1}$ by $[1/(n_1 - p - 1)]^{-1}$. Hence we define an alternative estimator for ς to be

$$\hat{\varsigma}^{S1} = A^{-1}\Phi^{S1}A$$

where the j'th diagonal element of the diagonal matrix Φ^{S1} is $\phi_j^{S1} = (n_1 - p - 1)\varphi_j^{ST}$. It is clear that the natural ordering of the ϕ_i^{S1}'s is preserved.

METHOD II.

For $1 \leq j \leq p$, let the φ_j^{ST}'s be defined as in Method I and $\varrho_j = 1/\beta_j$. The natural ordering of the ϱ_j's, that is $\varrho_1 \geq \cdots \geq \varrho_p \geq 0$, may not be satisfied. Applying Stein's isotonic regression to the ϱ_j's, we get a new set of ϱ_j's, denoted by ϱ_j^{ST}, $1 \le j \le p$, such that $\varrho_1^{ST} \ge \cdots \ge \varrho_p^{ST} \ge 0$. We now consider $\phi_1^{ST} = \varphi_j^{ST}/\varrho_j^{ST}$, $1 \le j \le p$. Again, the natural ordering on the ϕ_j^{ST}'s is given by $\phi_1^{ST} \ge \cdots \ge \phi_p^{ST} \ge 0$. To preserve this ordering, we apply Stein's isotonic regression to $\varphi_j^{ST}/\varrho_j^{ST}$ to get a new set of ϕ_j^{ST}'s, denoted by ϕ_j^{S2}, $1 \le j \le p$, such that $\phi_1^{S2} \ge \cdots \phi_p^{S2} \ge 0$ is satisfied. Finally we define an alternative estimator for ς to be

$$\hat{\varsigma}^{S2} = A^{-1}\Phi^{S2}A$$

where the j'th diagonal element of the diagonal matrix Φ^{S2} is ϕ_j^{S2}.

8 Monte Carlo Study

From the rather complicated construction of the Stein-type estimators, we observe that an analytical treatment of the risk performance of these estimators is not possible at this point. In this section we shall use Monte Carlo simulations to study

1. the risk performances, under L_1 loss, of the alternative estimators for ς which we have constructed in previous sections.

2. the risk performances of suitably scaled versions of these estimators under L_2 loss.

For the simulations, independent standard normal variates are generated by the IMSL subroutine DRNNOA and the eigenvalue decomposition uses the IMSL subroutine DEVCSF.

UNDER L_1 LOSS

For this study, we take $p = 10$, $n_1 = 12, 25$ and $n_2 = 12, 25$. Tables 1 to 4 give the average losses and their estimated standard deviations of the estimators: $\hat{\varsigma}^{BU}$, $\hat{\varsigma}^{AU}$, $\hat{\varsigma}^{BE}$, $\hat{\varsigma}^{S1}$ and $\hat{\varsigma}^{S2}$ based on 500 independent replications. As it is, the estimator $\hat{\varsigma}^{BE}$ is not well-defined. In this study, we choose

$$\begin{aligned} b &= 25000, \\ c &= 2(p^2+p-4)(n_2-p+3)\sqrt{b}/[\sqrt{p}(n_1-p-1)(n_2-p+7)]. \end{aligned}$$

These values of b and c are chosen to ensure good risk performance of $\hat{\varsigma}^{BE}$ at $n_1 = 12$, $n_2 = 12$ when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are equal.

We shall now summarize the results of this numerical study.

1. Under L_1 loss, the average losses of the estimators $\hat{\varsigma}^{AU}$, $\hat{\varsigma}^{BE}$, $\hat{\varsigma}^{S1}$ and $\hat{\varsigma}^{S2}$ compare favorably with that of $\hat{\varsigma}^{BU}$. This is most significant when the eigenvalues of ς are close together.

2. Among the alternative estimators, $\hat{\varsigma}^{S2}$ gives the most substantial savings in risk when the eigenvalues of ς are equal.

3. As a check on the accuracy of the Monte Carlo study, we observe that as predicted by theory, the following order holds: average loss of $\hat{\varsigma}^{BU}$ $\geq$ average loss of $\hat{\varsigma}^{AU}$ $\geq$ average loss of $\hat{\varsigma}^{BE}$. Also we note that the risk of $\hat{\varsigma}^{BU}$ is known exactly and is given in Theorem 5. We observe that the simulated average losses of $\hat{\varsigma}^{BU}$ agree with the theoretical values.

UNDER L_2 LOSS

For this study, we take $p = 10$, $n_1 = 14, 25$ and $n_2 = 14, 25$. From the forms of the best usual estimator $\hat{\varsigma}^{BU}$ under L_1 loss and the best usual estimator $\check{\varsigma}^{BU}$ under L_2 loss, we observe that a natural scaling factor would be $c = (n_1 - p)(n_1 - p - 3)/[(n_1 - 1)(n_1 - p - 1)]$. Hence we define our alternative estimators for ς to be

$$\check{\varsigma}^{XY} = c\hat{\varsigma}^{XY}$$

where XY represents one of the following: AU, BE, $S1$, $S2$. Tables 5 to 8 give the average losses and their estimated standard deviations of the estimators $\check{\varsigma}^{BU}$, $\check{\varsigma}^{AU}$, $\check{\varsigma}^{BE}$, $\check{\varsigma}^{S1}$ and $\check{\varsigma}^{S2}$ based on 500 independent replications. The results of this simulation study are summarized as follows:

1. This study indicates that the risk performances of these estimators are similar to that of their L_1 counterparts. However the effects are somewhat lessened.

2. Among the alternative estimators of ς, $\check{\varsigma}^{S2}$ does best when the eigenvalues of ς are equal.

3. We observe that in every instance the following order holds: average loss of $\check{\varsigma}^{BU}$ $\geq$ average loss of $\check{\varsigma}^{AU}$ $\geq$ average loss of $\check{\varsigma}^{BE}$. This suggests a strong possibility that $\check{\varsigma}^{BE}$ dominates $\check{\varsigma}^{AU}$ and $\check{\varsigma}^{AU}$ dominates $\check{\varsigma}^{BU}$.

4. As a check on the accuracy of the Monte Carlo simulations, the risk of $\check{\varsigma}^{BU}$ is known exactly and is given in Theorem 6. We observe that the simulated risks of $\check{\varsigma}^{BU}$ agree with the theoretical values.

Finally we wish to remark that in our simulation, 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 Tables 1 to 8) is probably a pessimistic measure of the variability of the relative magnitude of the average losses.

9 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. Special thanks go to Professors Ingram Olkin and Iain Johnstone for reading my dissertation and giving me several enlightening suggestions. Finally I would like to thank Professor Robb Muirhead for some very constructive comments when I visited the University of Michigan.

TABLE 1

$n_1 = 12 \qquad n_2 = 12$

Average losses of estimators for the estimation of $\Sigma_2\Sigma_1^{-1}$ *under* L_1 *loss (Estimated standard errors are in parenthesis)*

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\varsigma}^{BU}$	$\hat{\varsigma}^{AU}$	$\hat{\varsigma}^{BE}$	$\hat{\varsigma}^{S1}$	$\hat{\varsigma}^{S2}$
(1,1,1,1,1,	11.33	10.94	9.28	10.86	6.36
1,1,1,1,1)	(0.07)	(0.07)	(0.08)	(0.07)	(0.11)
(10,10,10,10,10,	11.26	11.10	11.07	11.08	9.56
0.1,0.1,0.1,0.1,0.1)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
(25,25,25,25,25,	11.34	11.05	11.02	11.00	7.99
25,25,25,0.1,0.1)	(0.08)	(0.08)	(0.08)	(0.08)	(0.10)
(30,30,30,0.1,0.1,	11.37	11.30	11.29	11.28	10.58
0.1,0.1,0.1,0.1,0.1)	(0.14)	(0.14)	(0.14)	(0.14)	(0.13)
(50,0.1,0.1,0.1,0.1,	11.44	11.43	11.42	11.43	11.36
0.1,0.1,0.1,0.1,0.1)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)
(20,20,20,5,5,	11.34	11.12	10.72	11.10	8.69
5,5,1,1,1)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
(100,90,80,70,60,	11.30	10.98	10.89	10.93	7.25
50,40,30,20,10)	(0.08)	(0.08)	(0.08)	(0.08)	(0.10)
(512,256,128,64,32,	11.34	11.23	11.19	11.23	10.14
16,8,4,2,1)	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)
(10^9,10^8,10^7,10^6,10^5,	11.42	11.40	11.40	11.41	11.34
10^4,10^3,10^2,10^1,10^0)	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)

TABLE 2

$n_1 = 25 \qquad n_2 = 25$

Average losses of estimators for the estimation of $\Sigma_2\Sigma_1^{-1}$ under L_1 loss (Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\varsigma}^{BU}$	$\hat{\varsigma}^{AU}$	$\hat{\varsigma}^{BE}$	$\hat{\varsigma}^{S1}$	$\hat{\varsigma}^{S2}$
(1,1,1,1,1,	15.34	12.47	11.86	10.48	4.58
1,1,1,1,1)	(0.11)	(0.11)	(0.11)	(0.11)	(0.14)
(10,10,10,10,10,	15.33	14.36	14.27	13.58	11.97
0.1,0.1,0.1,0.1,0.1)	(0.13)	(0.13)	(0.13)	(0.13)	(0.14)
(25,25,25,25,25,	15.28	13.30	13.27	11.85	7.93
25,25,25,0.1,0.1)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)
(30,30,30,0.1,0.1,	15.23	14.81	14.76	14.41	14.15
0.1,0.1,0.1,0.1,0.1)	(0.17)	(0.16)	(0.16)	(0.16)	(0.17)
(50,0.1,0.1,0.1,0.1,	15.11	15.05	14.98	15.02	15.48
0.1,0.1,0.1,0.1,0.1)	(0.27)	(0.27)	(0.27)	(0.27)	(0.29)
(20,20,20,5,5,	15.30	14.05	13.98	13.70	12.06
5,5,1,1,1)	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)
(100,90,80,70,60,	15.32	13.21	13.20	12.17	8.52
50,40,30,20,10)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
(512,256,128,64,32,	15.26	14.77	14.76	14.78	14.49
16,8,4,2,1)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
(10^9,10^8,10^7,10^6,10^5,	15.12	15.09	15.09	15.12	15.16
10^4,10^3,10^2,10^1,10^0)	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)

TABLE 3

$n_1 = 12 \qquad n_2 = 25$

Average losses of estimators for the estimation of $\Sigma_2\Sigma_1^{-1}$ *under* L_1 *loss (Estimated standard errors are in parenthesis)*

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\varsigma}^{BU}$	$\hat{\varsigma}^{AU}$	$\hat{\varsigma}^{BE}$	$\hat{\varsigma}^{S1}$	$\hat{\varsigma}^{S2}$
(1,1,1,1,1,	11.17	10.82	5.71	10.75	5.01
1,1,1,1,1)	(0.07)	(0.07)	(0.05)	(0.07)	(0.12)
(10,10,10,10,10,	11.12	10.97	10.82	10.95	8.87
0.1,0.1,0.1,0.1,0.1)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
(25,25,25,25,25,	11.16	10.90	10.79	10.85	6.97
25,25,25,0.1,0.1)	(0.08)	(0.08)	(0.08)	(0.08)	(0.11)
(30,30,30,0.1,0.1,	11.21	11.15	11.09	11.13	10.06
0.1,0.1,0.1,0.1,0.1)	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)
(50,0.1,0.1,0.1,0.1,	11.45	11.44	11.37	11.44	11.37
0.1,0.1,0.1,0.1,0.1)	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)
(20,20,20,5,5,	11.20	11.00	10.18	11.00	8.05
5,5,1,1,1)	(0.10)	(0.10)	(0.09)	(0.10)	(0.10)
(100,90,80,70,60,	11.17	10.88	10.74	10.85	6.21
50,40,30,20,10)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
(512,256,128,64,32,	11.25	11.15	11.08	11.16	9.73
16,8,4,2,1)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
(10^9,10^8,10^7,10^6,10^5,	11.41	11.40	11.40	11.40	11.31
10^4,10^3,10^2,10^1,10^0)	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)

TABLE 4

$n_1 = 25 \qquad n_2 = 12$

Average losses of estimators for the estimation of $\Sigma_2\Sigma_1^{-1}$ *under* L_1 *loss (Estimated standard errors are in parenthesis)*

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\varsigma}^{BU}$	$\hat{\varsigma}^{AU}$	$\hat{\varsigma}^{BE}$	$\hat{\varsigma}^{S1}$	$\hat{\varsigma}^{S2}$
(1,1,1,1,1,	17.67	14.46	14.01	12.68	8.15
1,1,1,1,1)	(0.10)	(0.11)	(0.10)	(0.12)	(0.13)
(10,10,10,10,10,	17.72	16.64	16.60	16.06	15.00
0.1,0.1,0.1,0.1,0.1)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
(25,25,25,25,25,	17.63	15.45	15.44	14.30	11.61
25,25,25,0.1,0.1)	(0.11)	(0.12)	(0.12)	(0.12)	(0.13)
(30,30,30,0.1,0.1,	17.66	17.19	17.18	16.89	16.68
0.1,0.1,0.1,0.1,0.1)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
(50,0.1,0.1,0.1,0.1,	17.75	17.69	17.67	17.67	17.88
0.1,0.1,0.1,0.1,0.1)	(0.27)	(0.27)	(0.27)	(0.27)	(0.28)
(20,20,20,5,5,	17.66	16.13	16.07	15.73	14.38
5,5,1,1,1)	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)
(100,90,80,70,60,	17.67	15.19	15.18	14.11	11.15
50,40,30,20,10)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
(512,256,128,64,32,	17.70	17.07	17.07	17.05	16.70
16,8,4,2,1)	(0.17)	(0.16)	(0.16)	(0.16)	(0.16)
(10^9,10^8,10^7,10^6,10^5,	17.74	17.71	17.71	17.74	17.77
10^4,10^3,10^2,10^1,10^0)	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)

TABLE 5

$n_1 = 14 \qquad n_2 = 14$

Average losses of estimators for the estimation of $\Sigma_2\Sigma_1^{-1}$ under L_2 loss (Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	ζ^{BU}	ζ^{AU}	ζ^{BE}	ζ^{S1}	ζ^{S2}
(1,1,1,1,1,	0.923	0.912	0.898	0.915	0.894
1,1,1,1,1)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(10,10,10,10,10,	0.924	0.918	0.917	0.920	0.915
0.1,0.1,0.1,0.1,0.1)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
(25,25,25,25,25,	0.923	0.914	0.913	0.916	0.904
25,25,25,0.1,0.1)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(30,30,30,0.1,0.1,	0.920	0.917	0.917	0.918	0.918
0.1,0.1,0.1,0.1,0.1)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(50,0.1,0.1,0.1,0.1,	0.925	0.924	0.924	0.924	0.929
0.1,0.1,0.1,0.1,0.1)	(0.005)	(0.005)	(0.005)	(0.005)	(0.003)
(20,20,20,5,5,	0.921	0.914	0.912	0.917	0.904
5,5,1,1,1)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(100,90,80,70,60,	0.923	0.913	0.913	0.916	0.898
50,40,30,20,10)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(512,256,128,64,32,	0.922	0.918	0.918	0.920	0.916
16,8,4,2,1)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
(10^9,10^8,10^7,10^6,10^5,	0.924	0.923	0.923	0.924	0.923
10^4,10^3,10^2,10^1,10^0)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)

TABLE 6

$n_1 = 25 \qquad n_2 = 25$

Average losses of estimators for the estimation of $\Sigma_2\Sigma_1^{-1}$ under L_2 loss (Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	ξ^{BU}	ξ^{AU}	ξ^{BE}	ξ^{S1}	ξ^{S2}
(1,1,1,1,1,	0.623	0.548	0.533	0.515	0.408
1,1,1,1,1)	(0.008)	(0.008)	(0.008)	(0.007)	(0.005)
(10,10,10,10,10,	0.625	0.595	0.593	0.578	0.551
0.1,0.1,0.1,0.1,0.1)	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)
(25,25,25,25,25,	0.621	0.566	0.565	0.540	0.468
25,25,25,0.1,0.1)	(0.008)	(0.008)	(0.008)	(0.007)	(0.006)
(30,30,30,0.1,0.1,	0.624	0.611	0.610	0.601	0.604
0.1,0.1,0.1,0.1,0.1)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
(50,0.1,0.1,0.1,0.1,	0.629	0.628	0.626	0.627	0.650
0.1,0.1,0.1,0.1,0.1)	(0.010)	(0.010)	(0.010)	(0.010)	(0.007)
(20,20,20,5,5,	0.625	0.588	0.586	0.582	0.544
5,5,1,1,1)	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)
(100,90,80,70,60,	0.623	0.565	0.565	0.547	0.472
50,40,30,20,10)	(0.008)	(0.008)	(0.008)	(0.008)	(0.005)
(512,256,128,64,32,	0.626	0.611	0.611	0.612	0.604
16,8,4,2,1)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
(10^9,10^8,10^7,10^6,10^5,	0.627	0.626	0.626	0.627	0.628
10^4,10^3,10^2,10^1,10^0)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)

TABLE 7

$n_1 = 14 \qquad n_2 = 25$

Average losses of estimators for the estimation of $\Sigma_2\Sigma_1^{-1}$ *under* L_2 *loss (Estimated standard errors are in parenthesis)*

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	ζ^{BU}	ζ^{AU}	ζ^{BE}	ζ^{S1}	ζ^{S2}
(1,1,1,1,1,	0.919	0.910	0.889	0.913	0.898
1,1,1,1,1)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
(10,10,10,10,10,	0.912	0.907	0.905	0.909	0.907
0.1,0.1,0.1,0.1,0.1)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
(25,25,25,25,25,	0.917	0.909	0.909	0.912	0.903
25,25,25,1,1)	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)
(30,30,30,0.1,0.1,	0.912	0.909	0.909	0.910	0.915
0.1,0.1,0.1,0.1,0.1)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
(50,0.1,0.1,0.1,0.1,	0.915	0.915	0.914	0.915	0.928
0.1,0.1,0.1,0.1,0.1)	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)
(20,20,20,5,5,	0.918	0.911	0.909	0.914	0.901
5,5,1,1,1)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
(100,90,80,70,60,	0.917	0.909	0.909	0.912	0.896
50,40,30,20,10)	(0.012)	(0.012)	(0.012)	(0.011)	(0.010)
(512,256,128,64,32,	0.914	0.910	0.910	0.912	0.907
16,8,4,2,1)	(0.009)	(0.009)	(0.009)	(0.009)	(0.006)
(10^9,10^8,10^7,10^6,10^5,	0.913	0.913	0.913	0.913	0.913
10^4,10^3,10^2,10^1,10^0)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)

TABLE 8

$n_1 = 25 \qquad n_2 = 14$

Average losses of estimators for the estimation of $\Sigma_2\Sigma_1^{-1}$ under L_2 loss (Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	ζ^{BU}	ζ^{AU}	ζ^{BE}	ζ^{S1}	ζ^{S2}
(1,1,1,1,1,	0.691	0.603	0.590	0.567	0.473
1,1,1,1,1)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
(10,10,10,10,10,	0.689	0.655	0.654	0.641	0.620
0.1,0.1,0.1,0.1,0.1)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
(25,25,25,25,25,	0.691	0.627	0.627	0.601	0.541
25,25,25,0.1,0.1)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
(30,30,30,0.1,0.1,	0.687	0.672	0.671	0.663	0.661
0.1,0.1,0.1,0.1,0.1)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)
(50,0.1,0.1,0.1,0.1,	0.687	0.685	0.684	0.685	0.704
0.1,0.1,0.1,0.1,0.1)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
(20,20,20,5,5,	0.689	0.645	0.643	0.638	0.601
5,5,1,1,1)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
(100,90,80,70,60,	0.690	0.621	0.621	0.600	0.531
50,40,30,20,10)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
(512,256,128,64,32,	0.687	0.668	0.667	0.668	0.659
16,8,4,2,1)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
(10^9,10^8,10^7,10^6,10^5,	0.686	0.685	0.685	0.686	0.687
10^4,10^3,10^2,10^1,10^0)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)

References

[1] BERGER, J. O. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.* **8** 545-571.

[2] DASGUPTA, A. (1986). A nonparametric theorem on decision theory for nonnegative functionals: With applications. Technical Report No. 86-25, Department of Statistics, Purdue University.

[3] DEY, D. K. (1986). Simultaneous estimation of eigenvalues. Unpublished manuscript.

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

[5] HAFF, L. R. (1979a). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.* **7** 1264-1276.

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

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

[8] LOH, W. L. (1988). Estimating covariance matrices I. Technical report, Department of Statistics, Purdue University, West Lafayette.

[9] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

[10] MUIRHEAD, R. J. (1987). Developments in eigenvalue estimation. *Advances in Multivariate Statistical Analysis*, 277-288, ed. A. K. Gupta. D. Reidel, Boston.

[11] MUIRHEAD, R. J. and LEUNG, P. L. (1988). On estimating characteristic roots in a two-sample problem. *Statistical Decision Theory and Related Topics IV*, Vol. 1, 397-402, ed. S. S. Gupta & J. O. Berger. Springer Verlag, New York.

[12] MUIRHEAD, R. J. and VERATHAWORN, T. (1985). On estimating the latent roots of $\Sigma_1\Sigma_2^{-1}$. *Multivariate Analy.* **6**, 431-447, ed. P. R. Krishnaiah, North Holland, Amsterdam.

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

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

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

[16] STEIN, C. (1977b). Lectures on the theory of estimation of many parameters. (In Russian.) *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.

[17] STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135-1151.