ESTIMATING COVARIANCE MATRICES I

by

Wei–Liem Loh
Purdue University

Technical Report #88-37

Department of Statistics
Purdue University

August 1988

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.

*

ESTIMATING COVARIANCE MATRICES I

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 $L(\hat{\Sigma}_1, \hat{\Sigma}_2; \Sigma_1, \Sigma_2) = \sum_i \{\mathrm{tr}(\Sigma_i^{-1}\hat{\Sigma}_i) - \log \mid \Sigma_i^{-1}\hat{\Sigma}_i \mid -p\}$. 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.

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), (1977a), (1977b), Haff (1980), (1982), (1988) and Dey and Srinivasan (1985), (1986). In this paper, its 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 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 $\mathrm{tr}A$ 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 $A = \mathrm{diag}(a_{11}, \ldots, a_{pp})$. 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:

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

This loss function is convex and is the natural extension of Stein's loss in the one sample case, where Stein's loss is given by

$$L_S\hat{\Sigma}; \Sigma) = \mathrm{tr}(\Sigma^{-1}\hat{\Sigma}) - \log \mid \Sigma^{-1}\hat{\Sigma} \mid -p.$$

The loss function L_S was first considered by Stein (1956). As it is, the above problem is a canonical formulation of the following more common situation: Let $X_1, \ldots, X_{n_1+1}$ and $Y_1, \ldots, Y_{n_2+1}$ be two samples from two multivariate normal distributions $N_p(\xi_1, \Sigma_1)$ and $N_p(\xi_2, \Sigma_2)$ respectively. We wish to estimate (Σ_1, Σ_2). The sufficient statistics are

$$\begin{aligned} \bar{X} &= [1/(n_1+1)] \sum_i X_i, \qquad S_1 = \sum_i (X_i - \bar{X})^2, \\ \bar{Y} &= [1/(n_2+1)] \sum_i Y_i, \qquad S_2 = \sum_i (Y_i - \bar{Y})^2. \end{aligned}$$

Since $S_1 \sim W_p(\Sigma_1, n_1)$ and $S_2 \sim W_p(\Sigma_2, n_2)$ with S_1, S_2 independent, this reduces to the above formulation.

In the course of giving a talk at Stanford, I was informed that Bilodeau (1987) has also worked on the estimation of more than one covariance matrix. However the method used, loss function and results are distinctly different from ours.

2 Two Important Identities

We shall first state and prove two important identities: namely the Normal and Wishart identities. These identities are crucial in the developments that follow.

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.

The following lemma is essentally taken from Stein (1981) and hence the proof is omitted.

Lemma 1 (Stein's Lemma) *Let* $Y = (Y_1, \ldots, Y_p)' \sim N_p(\xi, I)$ *and* $h : R^p \to R$ *be an almost differentiable function with* $E[\sum_i \mid \nabla_i h(Y) \mid]$ *finite. Then*

$$E\nabla h(Y) = E[(Y - \xi)h(Y)],$$

where $\nabla = (\nabla_1, \ldots, \nabla_p)'$ *with* $\nabla_i = \partial/\partial Y_i$.

Now we state and prove the Normal identity. The Normal identity was first proved by Stein (1973).

Theorem 1 (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 $\nabla = (\partial/\partial X_1, \ldots, \partial/\partial X_p)'$.

PROOF. Let $Y = \Sigma^{-1/2}X$ and $h(Y) = g(X)$. Then $Y \sim N_p(\Sigma^{-1/2}\xi, I)$. We observe that for all $i, k = 1, \ldots, p$, $E \mid \partial h_k(Y)/\partial Y_i \mid$ is finite and that h_k is almost differentiable. Now it follows from Stein's lemma that for each $i, k = 1, \ldots, p$,

$$\begin{aligned} Eh_k(Y)(Y - \Sigma^{-1/2}\xi)_i &= E\frac{\partial h_k(Y)}{\partial Y_i} \\ &= E\sum_j \frac{\partial g_k(X)}{\partial X_j}\frac{\partial X_j}{\partial Y_i} \\ &= E\sum_j (\Sigma^{1/2})_{ij}\frac{\partial g_k(X)}{\partial X_j}. \end{aligned}$$

This implies that

$$E[\Sigma^{-1/2}(X - \xi)g'(X)]_{ik} = E[\Sigma^{1/2}\nabla g'(X)]_{ik}.$$

Thus we conclude that

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

This completes the proof. □

Let S_p denote the set of $p \times p$ positive definite matrices. Also we write for $1 \leq i, j \leq p$,

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

where δ_{ij} denotes the Kronecker delta.

The following theorem and its proof are taken from the unpublished notes of Stein (1977a).

Theorem 2 (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 $n \geq p$. 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*

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

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

PROOF. We first consider the special case $\Sigma = I$. We recall from Stein's lemma that

$$EX_{ik}f(X) = E\partial f(X)/\partial X_{ik}$$

for almost differentiable $f : R^{p\times n} \to R$ provided the expectation on the r.h.s. exists. If we specialize f to $x \mapsto x_{jk}h_{ij}(xx')$ with $h_{ij} : \mathcal{S}_p \to R$ almost differentiable, and such that the expectations used below exist, then

$$\begin{aligned} EX_{ik}X_{jk}h_{ij}(S) &= E\frac{\partial}{\partial X_{ik}}[X_{jk}h_{ij}(S)] \\ &= E[\delta_{ij}h_{ij}(S) + X_{jk}\frac{\partial}{\partial X_{ik}}h_{ij}(S)] \\ &= E[\delta_{ij}h_{ij}(S) + X_{jk}\sum_{j'=1}^{p}\frac{\partial h_{ij}(S)}{\partial S_{ij'}}(1+\delta_{ij'})X_{j'k}] \\ &= E[\delta_{ij}h_{ij}(S) + 2X_{jk}\sum_{j'=1}^{p}\tilde{\nabla}_{j'i}h_{ij}(S)X_{j'k}]. \end{aligned}$$

Summing over k from 1 to n, we get

$$ES_{ji}h_{ij}(S) = E[n\delta_{ij}h_{ij}(S) + 2(S\tilde{\nabla})_{ji}h_{ij}(S)]. \tag{3}$$

If we consider the $h_{ij} : \mathcal{S}_p \to R$ as coordinates of a function $h : \mathcal{S}_p \to R^{p\times p}$ we find,by summing (3) over i and j from 1 to p and letting $h(s) = g(s)s^{-1}$ that

$$E\mathrm{tr}g(S) = E\mathrm{tr}[nS^{-1}g(S) + 2S\tilde{\nabla}(g(S)S^{-1})], \tag{4}$$

which is the special case $\Sigma = I$ of (2). To prove (2) when Σ is arbitrary, let

$$\Sigma = \alpha\alpha'.$$

With S and X as in the first part of the proof, that is, with the columns of X independently normally distributed with mean 0 and the identity as covariance matrix, and with $S = XX'$, let

$$S^{(1)} = \alpha S \alpha', \qquad s^{(1)} = \alpha s \alpha'$$

and also

$$\tilde{\nabla}^{(1)}_{ij} = (1/2)(1+\delta_{ij})\partial/\partial s^{(1)}_{ij}.$$

It is not difficult to verify that

$$\tilde{\nabla} = \alpha' \tilde{\nabla}^{(1)} \alpha.$$

With

$$h(s^{(1)}) = \alpha g(s) \alpha',$$

we have

(5) $$\mathrm{tr} g(s) = \mathrm{tr} \Sigma^{-1} h(s^{(1)})$$

and

(6) $$\mathrm{tr} s^{-1} g(s) = \mathrm{tr} s^{(1)^{-1}} h(s^{(1)}),$$

and

$$\begin{aligned} & \mathrm{tr} s \tilde{\nabla}(g(s) s^{-1}) \\ = \ & \mathrm{tr} \alpha^{-1} s^{(1)} \alpha^{-1'} \alpha' \tilde{\nabla}^{(1)} \alpha [\alpha^{-1} h(s^{(1)}) \alpha^{-1'} (\alpha^{-1} s^{(1)} \alpha^{-1'})^{-1}] \\ = \ & \mathrm{tr} s^{(1)} \tilde{\nabla}^{(1)} [h(s^{(1)}) s^{(1)^{-1}}]. \qquad (7) \end{aligned}$$

Substituting (5), (6) and (7) in (4), we find

$$E \mathrm{tr} \Sigma^{-1} h(S^{(1)}) = E \mathrm{tr}[n S^{(1)^{-1}} h(S^{(1)}) + 2 S^{(1)} \tilde{\nabla}^{(1)} (h(S^{(1)}) S^{(1)^{-1}})]$$

which is (2) except for the difference in notation. □

The Wishart identity was proved by Stein (1975) and Haff (1977) independently. Next we state two lemmas which are proved in Haff (1979b) and (1981) respectively.

Lemma 2 *The matrix*

$$\begin{aligned} \partial S^{-1}/\partial s_{kl} &= -S^{-1} e_k e'_k S^{-1} \qquad k = l \\ &= -S^{-1}(e_k e'_l + e_l e'_k) S^{-1} \qquad k \neq l, \end{aligned}$$

where e_i denotes the ith unit vector.

Lemma 3 *Let F and G be matrix functions of S. Assuming all relevant products and derivatives exist, we have*

$$\tilde{\nabla}(FG) = (F'\tilde{\nabla})'G + (\tilde{\nabla}F)G.$$

With these two lemmas in hand , we prove the following corollary of the Wishart identity.

Corollary 1 *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 $n \geq p$. Let $g : S_p \to R^{p\times p}$ 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.

PROOF. First we observe from Lemma 2 that

$$\begin{aligned}
& \mathrm{tr}S(g'(S)\tilde{\nabla})'S^{-1} \\
= & \sum_{i,k,l} s_{ik}(g'(S)\tilde{\nabla})_{lk}s^{li} \\
= & \sum_{i,j,k,l} g_{jl}(\tilde{\nabla}_{jk}s^{li})s_{ik} \\
= & -\sum_{i,j,k,l} \{g_{jl}[S^{-1}e_j e'_j S^{-1}]_{li}s_{ik}\delta_{jk} \\
& +(1/2)g_{jl}[S^{-1}(e_j e'_k + e_k e'_j)S^{-1}]_{li}s_{ik}(1-\delta_{jk})\} \\
= & -(p+1)\mathrm{tr}S^{-1}g(S)/2,
\end{aligned}$$

where $S^{-1} = (s^{il})$. This implies that

$$\begin{aligned}
\mathrm{tr}S\tilde{\nabla}(g(S)S^{-1}) &= \mathrm{tr}S[(\tilde{\nabla}g(S))S^{-1} + (g'(S)\tilde{\nabla})'(S^{-1})] \\
&= \mathrm{tr}[\tilde{\nabla}g(S) + S(g'(S)\tilde{\nabla})'(S^{-1})] \\
&= \mathrm{tr}[\tilde{\nabla}g(S) - (p+1)S^{-1}g(S)/2]. \qquad (8)
\end{aligned}$$

Substituting (8) into (2), we get

$$E\mathrm{tr}\Sigma^{-1}g(S) = E\mathrm{tr}[(n-p-1)S^{-1}g(S) + 2\tilde{\nabla}g(S)].$$

This completes the proof. □

3 Equivariant Estimators

The problem which we are concerned with is invariant under the following group of transformations:

$$(9)\ \Sigma_i \to A\Sigma_i A', \qquad S_i \to AS_iA' \qquad \forall A \in GL(p,R), \qquad i=1,2.$$

Theorem 3 *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 (9), $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ is an equivariant estimator of (Σ_1, Σ_2) if and only if $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ can be expressed as*

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

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 $F = diag(f_1, \ldots, f_p)$.

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

$$\begin{aligned} \hat{\Sigma}_i(S_1,S_2,n_1,n_2) &= A^{-1}\hat{\Sigma}_i(AS_1A', AS_2A', n_1, n_2)A'^{-1}, \\ (10) \qquad & \qquad \forall A \in GL(p,R). \end{aligned}$$

We observe that $\exists B \in GL(p,R)$ such that $B(S_1+S_2)B' = I$ and $BS_2B' = F$, where $F = \text{diag}(f_1,\ldots,f_p)$ with $f_1 \geq \cdots \geq f_p$. Hence it follows from (10) 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

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

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

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

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

4 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 $\exists B \in GL(p, R)$ such that $BS_1B' = I - F$ and $BS_2B' = F$ where $F = \mathrm{diag}(f_1, \ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$. In this section, the partial derivatives of B^{-1} and F with respect to S_1 and S_2 are computed.

Theorem 4 *Let $S_1 \sim W_p(\Sigma_1, n_1)$ and $S_2 \sim W_p(\Sigma_2, n_2)$. Then with F, $B = (b_{il})$, $B^{-1} = (b^{il})$ as defined above, we have*

$$\begin{aligned}
\tilde{\nabla}_{jk}^{(1)} f_i &= -f_i b_{ij} b_{ik}, \\
\tilde{\nabla}_{jk}^{(2)} f_i &= (1-f_i) b_{ij} b_{ik}, \\
\tilde{\nabla}_{jk}^{(1)} b^{il} &= \frac{1}{2}[\sum_{i' \neq l} b^{ii'}(b_{i'j}b_{lk} + b_{i'k}b_{lj})\frac{f_{i'}}{f_{i'} - f_l}] + \frac{1}{2} b^{il} b_{lj} b_{lk}, \\
\tilde{\nabla}_{jk}^{(2)} b^{il} &= \frac{1}{2}[\sum_{i' \neq l} b^{ii'}(b_{i'j}b_{lk} + b_{i'k}b_{lj})\frac{1 - f_{i'}}{f_l - f_{i'}}] + \frac{1}{2} b^{il} b_{lj} b_{lk}.
\end{aligned}$$

PROOF. On differentiating $S_1 = B^{-1}(I-F)B'^{-1}$ and $S_2 = B^{-1}FB'^{-1}$, we have

$$\begin{aligned}
\mathrm{d}S_1 &= (\mathrm{d}B^{-1})(I-F)B'^{-1} - B^{-1}(\mathrm{d}F)B'^{-1} + B^{-1}(I-F)(\mathrm{d}B'^{-1}), \\
\mathrm{d}S_2 &= (\mathrm{d}B^{-1})FB'^{-1} + B^{-1}(\mathrm{d}F)B'^{-1} + B^{-1}F(\mathrm{d}B'^{-1}).
\end{aligned}$$

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

$$\begin{aligned}
&(11) \quad B(\mathrm{d}S_1)B' = B(\mathrm{d}B^{-1})(I-F) - (\mathrm{d}F) + (I-F)(\mathrm{d}B'^{-1})B', \\
&(12) \quad B(\mathrm{d}S_2)B' = B(\mathrm{d}B^{-1})F + (\mathrm{d}F) + F(\mathrm{d}B'^{-1})B'.
\end{aligned}$$

It follows from (12) that

$$\begin{aligned}
(I-F)F^{-1}B(\mathrm{d}S_2)B' &= (I-F)F^{-1}B(\mathrm{d}B^{-1})F + (I-F)F^{-1}(\mathrm{d}F) \\
(13) \qquad &\quad +(I-F)(\mathrm{d}B'^{-1})B'.
\end{aligned}$$

CASE I. Suppose that $\mathrm{d}S_2 = 0$. Subtracting (11) from (13) gives

$$-B(\mathrm{d}S_1)B' = F^{-1}B(\mathrm{d}B^{-1})F - B(\mathrm{d}B^{-1}) + F^{-1}(\mathrm{d}F),$$

which reduces to

$$(14)\qquad (\mathrm{d}F) = -FB(\mathrm{d}S_1)B' - B(\mathrm{d}B^{-1})F + FB(\mathrm{d}B^{-1}).$$

Hence for all i, j, k,

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

Thus we conclude that

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

Next from (14) we have

$$(\mathrm{d}F)_{ij} = -f_i[B(\mathrm{d}S_1)B']_{ij} - [B(\mathrm{d}B^{-1})]_{ij} f_j + f_i[B(\mathrm{d}B^{-1})]_{ij}$$

which gives

$$[B(\mathrm{d}B^{-1})]_{ij}(f_i - f_j) = (\mathrm{d}F)_{ij} + f_i[B(\mathrm{d}S_1)B']_{ij}.$$

If $i \neq j$, then

$$\begin{aligned}[B(\mathrm{d}B^{-1})]_{ij} &= \frac{f_i}{f_i - f_j}[B(\mathrm{d}S_1)B']_{ij} \\ &= \frac{f_i}{f_i - f_j}\sum_{k,l} b_{ik}(\mathrm{d}s^{(1)}_{kl})b_{jl}.\end{aligned}$$

Furthermore, adding (11) to (12) gives

$$B(\mathrm{d}S_1)B' = B(\mathrm{d}B^{-1}) + (\mathrm{d}B'^{-1})B'.$$

Considering the diagonal elements, we get

$$\begin{aligned}[B(\mathrm{d}B^{-1})]_{ii} &= \frac{1}{2}[B(\mathrm{d}S_1)B']_{ii} \\ &= \frac{1}{2}\sum_{k,l} b_{ik}(\mathrm{d}s^{(1)}_{kl})b_{il}.\end{aligned}$$

Thus we have

$$\begin{aligned}(\mathrm{d}B^{-1})_{ij} &= \{B^{-1}[B(\mathrm{d}B^{-1})]\}_{ij} \\ &= \sum_{i'} b^{ii'}[B(\mathrm{d}B^{-1})]_{i'j} \\ &= \sum_{i'\neq j} b^{ii'}\frac{f_{i'}}{f_{i'} - f_j}\sum_{k,l} b_{i'k}(\mathrm{d}s^{(1)}_{kl})b_{jl} \;+\; \frac{1}{2}b^{ij}\sum_{k,l} b_{jk}(\mathrm{d}s^{(1)}_{kl})b_{jl}.\end{aligned}$$

Hence we conclude that

$$\tilde{\nabla}^{(1)}_{jk} b^{il} = \frac{1}{2}[\sum_{i' \neq l} b^{ii'}(b_{i'j}b_{lk} + b_{i'k}b_{lj})\frac{f_{i'}}{f_{i'} - f_l}] + \frac{1}{2}b^{il}b_{lj}b_{lk}.$$

CASE II. Suppose that $dS_1 = 0$. The proof of

$$\begin{aligned}\tilde{\nabla}^{(2)}_{jk} f_i &= (1 - f_i)b_{ij}b_{ik}, \\ \tilde{\nabla}^{(2)}_{jk} b^{il} &= -\frac{1}{2}[\sum_{i' \neq l} b^{ii'}(b_{i'j}b_{lk} + b_{i'k}b_{lj})\frac{1 - f_{i'}}{f_{i'} - f_l}] + \frac{1}{2}b^{il}b_{lj}b_{lk}.\end{aligned}$$

is similar to that of Case I and hence is omitted. This completes the proof. □

5 Unbiased Estimate of Risk

In this section we shall compute the unbiased estimate of the risk of an almost arbitrary equivariant estimator of (Σ_1, Σ_2). First we start with a lemma.

Lemma 4 *Let B, $\Phi = diag(\phi_1, \ldots, \phi_p)$ and $\Psi = diag(\psi_1, \ldots, \psi_p)$ be defined as in Theorem 3. Then*

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

PROOF.

$$\begin{aligned}& \mathrm{tr}\tilde{\nabla}^{(1)}(B^{-1}\Psi B'^{-1}) \\ =& \sum_{i,j} \tilde{\nabla}^{(1)}_{ij}(B^{-1}\Psi B'^{-1})_{ji} \\ =& \sum_{i,j,k} \tilde{\nabla}^{(1)}_{ij}(b^{jk}\psi_k b^{ik}) \\ =& \sum_{i,j,k} [(\tilde{\nabla}^{(1)}_{ij} b^{jk})\psi_k b^{ik} + b^{jk}(\tilde{\nabla}^{(1)}_{ij}\psi_k)b^{ik} + b^{jk}\psi_k(\tilde{\nabla}^{(1)}_{ij} b^{ik})].\end{aligned}$$

Now it follows from Theorem 4 that

$$\begin{aligned}
& \mathrm{tr}\tilde{\nabla}^{(1)}(B^{-1}\Psi B'^{-1}) \\
= & \sum_{i,j,k}\{\psi_k b^{ik}[\frac{1}{2}b^{jk}b_{ki}b_{kj} + \frac{1}{2}\sum_{i'\neq k} b^{ji'}(b_{i'i}b_{kj} + b_{i'j}b_{ki})\frac{f_{i'}}{f_{i'} - f_k}] \\
& +b^{jk}b^{ik}\sum_m(-f_m b_{mi}b_{mj})\frac{\partial\psi_k}{\partial f_m} \\
& +\psi_k b^{jk}[\frac{1}{2}b^{ik}b_{ki}b_{kj} + \frac{1}{2}\sum_{i'\neq k} b^{ii'}(b_{i'i}b_{kj} + b_{i'j}b_{ki})\frac{f_{i'}}{f_{i'} - f_k}]\} \\
= & \sum_k\{\psi_k[\frac{1}{2} + \frac{1}{2}\sum_{i'\neq k}\frac{f_{i'}}{f_{i'} - f_k}] + f_k\frac{\partial\psi_k}{\partial(1-f_k)} \\
& +\psi_k[\frac{1}{2} + \frac{1}{2}\sum_{i'\neq k}\frac{f_{i'}}{f_{i'} - f_k}]\} \\
= & \sum_k[\psi_k + f_k\frac{\partial\psi_k}{\partial(1-f_k)} + \psi_k\sum_{j\neq k}\frac{f_j}{f_j - f_k}].
\end{aligned}$$

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

With this lemma in hand, we shall now prove the main results of this section.

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

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

$\Psi = diag(\psi_1, \ldots, \psi_p)$, $B(S_1+S_2)B' = I$, $BS_2B' = F = diag(f_1, \ldots, f_p)$ 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

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

PROOF. We observe from the Wishart identity that

$$\begin{aligned} E\mathrm{tr}(\Sigma_1^{-1}\hat{\Sigma}_1) &= E\mathrm{tr}[2\tilde{\nabla}^{(1)}(\hat{\Sigma}_1) + (n_1-p-1)S_1^{-1}\hat{\Sigma}_1] \\ &= E\mathrm{tr}\{2\tilde{\nabla}^{(1)}[B^{-1}\Psi B'^{-1}] + (n_1-p-1)S_1^{-1}B^{-1}\Psi B'^{-1}\} \\ &= E\mathrm{tr}\{2\tilde{\nabla}^{(1)}[B^{-1}\Psi B'^{-1}] + (n_1-p-1)(I-F)^{-1}\Psi\}. \end{aligned}$$

Now it follows from Lemma 4 that

$$\begin{aligned} E\mathrm{tr}(\Sigma_1^{-1}\hat{\Sigma}_1) &= E\{\sum_i[\frac{n_1-p-1}{1-f_i}\psi_i + 2\psi_i\sum_{j\neq i}\frac{f_j}{f_j-f_i} \\ &\quad +2\psi_i + 2f_i\frac{\partial\psi_i}{\partial(1-f_i)}]\}. \end{aligned} \tag{15}$$

Finally the risk of $\hat{\Sigma}_1$ is given by

$$\begin{aligned} R_S(\hat{\Sigma}_1;\Sigma_1) &= EL_S(\hat{\Sigma}_1;\Sigma_1) \\ &= E[\mathrm{tr}(\Sigma_1^{-1}\hat{\Sigma}_1) - \log\mid\Sigma_1^{-1}\hat{\Sigma}_1\mid -p] \\ &= E[\mathrm{tr}(\Sigma_1^{-1}\hat{\Sigma}_1) - \log\mid S_1^{-1}\hat{\Sigma}_1\mid - \log\mid\Sigma_1^{-1}S_1\mid -p] \\ &= E[\mathrm{tr}(\Sigma_1^{-1}\hat{\Sigma}_1) - \log\mid S_1^{-1}\hat{\Sigma}_1\mid - \sum_i\log\chi^2_{n_1-i+1} - p] \\ &= E[\mathrm{tr}(\Sigma_1^{-1}\hat{\Sigma}_1) - \sum_i\log\frac{\psi_i}{1-f_i} - \sum_i\log\chi^2_{n_1-i+1} - p]. \end{aligned}$$

It follows now from (15) that

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

This completes the proof. □

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

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

$\Phi = diag(\phi_1,\ldots,\phi_p)$, $B(S_1+S_2)B' = I, BS_2B' = F = diag(f_1,\ldots,f_p)$ 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 $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ is given by

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

PROOF. The proof is analogous to that of Proposition 1 and hence is omitted. □

Theorem 5 *Let $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ be an estimator for (Σ_1, Σ_2) where*

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

$\Phi = diag(\phi_1, \ldots, \phi_p)$, $\Psi = diag(\psi_1, \ldots, \psi_p)$, $B(S_1 + S_2)B' = I, BS_2B' = F = diag(f_1, \ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$. Suppose Φ and Ψ both satisfy the conditions of the Wishart identity in the sense that

$$
\begin{aligned}
1.\ E\mathrm{tr}(\Sigma_1^{-1}\hat{\Sigma}_1) &= E\mathrm{tr}[2\tilde{\nabla}^{(1)}(\hat{\Sigma}_1) + (n_1 - p - 1)S_1^{-1}\hat{\Sigma}_1], \\
2.\ E\mathrm{tr}(\Sigma_2^{-1}\hat{\Sigma}_2) &= E\mathrm{tr}[2\tilde{\nabla}^{(2)}(\hat{\Sigma}_2) + (n_1 - p - 1)S_2^{-1}\hat{\Sigma}_2].
\end{aligned}
$$

Then under the loss function L given by (1), the risk of $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ is given by

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

PROOF. This theorem follows directly from Propositions 1 and 2 since the loss function under consideration is the sum of the respective loss functions of these two problems. □

6 Usual Estimators and Minimax Risk

The usual estimators $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ of (Σ_1, Σ_2) are of the form $(c_1 S_1, c_2 S_2)$ where c_1, c_2 are constants. The best usual estimator is that usual estimator which minimizes the risk among the usual estimators.

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, the best usual estimator $(\hat{\Sigma}_1^{BU}, \hat{\Sigma}_2^{BU})$ of (Σ_1, Σ_2) is $(S_1/n_1, S_2/n_2)$. Also,*

$$\begin{aligned} R(\hat{\Sigma}_1^{BU}, \hat{\Sigma}_2^{BU}; \Sigma_1, \Sigma_2) &= E[p \log n_1 + p \log n_2 \\ &\quad - \sum_i (\log \chi^2_{n_1 - i+1} + \log \chi^2_{n_2 - i+1})]. \end{aligned}$$

PROOF. The risk of estimators of the form $(c_1 S_1, c_2 S_2)$ is given by

$$\begin{aligned} R(c_1 S_1, c_2 S_2; \Sigma_1, \Sigma_2) &= E[\mathrm{tr}(\Sigma_1^{-1} c_1 S_1) + \mathrm{tr}(\Sigma_2^{-1} c_2 S_2) - \log \mid \Sigma_1^{-1} c_1 S_1 \mid \\ &\quad - \log \mid \Sigma_2^{-1} c_2 S_2 \mid -2p] \\ &= E[c_1 n_1 p + c_2 n_2 p - p \log c_1 - p \log c_2 \\ &\quad - \sum_i (\log \chi^2_{n_1 - i+1} + \log \chi^2_{n_2 - i+1}) - 2p]. \end{aligned}$$

This is minimized when $(c_1, c_2) = (1/n_1, 1/n_2)$. □

We shall now review very briefly some results concerning the minimax estimation of a single covariance matrix. Recall that Stein's loss is given by

$$L_S(\hat{\Sigma}; \Sigma) = \mathrm{tr}(\Sigma^{-1} \hat{\Sigma}) - \log \mid \Sigma^{-1} \hat{\Sigma} \mid -p.$$

The next result is due to Stein (1956).

Theorem 7 *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 nonsingularlower triangular, is $\hat{\Sigma}^{MM}(S) = TDT'$, where the j'th diagonal element of the diagonal matrix D is $1/(n+p-2j+1)$, $j = 1, \ldots, p$, and $S = TT'$, with T lower triangular. This estimator is minimax with risk*

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

We shall now give a two sample analogue of Theorem 7. To do so, we shall consider the class of equivariant estimators of (Σ_1, Σ_2) under the following group of transformations,

$$(16) \qquad \Sigma_i \to U_i \Sigma_i U_i', \qquad S_i \to U_i S_i U_i', \qquad i = 1, 2,$$

where U_i is a nonsingular lower triangular matrix.

Theorem 8 *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 (16) is $(\hat{\Sigma}_1^{MM}, \hat{\Sigma}_2^{MM}) = (T_1 D_1 T_1', T_2 D_2 T_2')$, where, for i=1,2, the j'th 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*

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

PROOF. The group of transformations considered here is solvable. Hence there exists a minimax estimator of (Σ_1, Σ_2) equivariant under the group of transformations given by (16). As in Stein (1956), $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ is equivariant if and only if

$$(17) \qquad (\hat{\Sigma}_1, \hat{\Sigma}_2) = (T_1 \Lambda_1 T_1', T_2 \Lambda_2 T_2'),$$

where, for i=1,2, $S_i = T_i T_i'$, with T_i lower triangular, and $\Lambda_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$. Since this group of transformations is transitive, without loss of generality, we shall consider only the case where $(\Sigma_1, \Sigma_2) = (I, I)$. For such a $(\hat{\Sigma}_1, \hat{\Sigma}_2)$, we have

$$\begin{aligned}
& R(\hat{\Sigma}_1, \hat{\Sigma}_2; I, I) \\
= \; & E\{\sum_{i=1}^{2}[\mathrm{tr}(\Lambda_i T_i' T_i) - \log | \Lambda_i | - \log | T_i T_i' | - p]\} \\
= \; & E \sum_{i=1}^{2}\{\sum_j (\lambda_{ij} \chi^2_{n_i + p - 2j + 1} - \log \lambda_{ij} - \log \chi^2_{n_i - j + 1}) - p\} \\
= \; & \sum_{i=1}^{2}\{\sum_j [(n_i + p - 2j + 1)\lambda_{ij} - \log \lambda_{ij} - E \log \chi^2_{n_i - j + 1}] - p\},
\end{aligned}$$

since the j'th diagonal element of $T_i' T_i$ is distributed as $\chi^2_{n_i + p - 2j + 1}$. This is minimized by

$$(18) \qquad \lambda_{ij} = 1/(n_i + p - 2j + 1), \qquad \forall i, j.$$

Thus a minimax estimator is given by (17) with the diagonal elements of Λ_i, $i = 1, 2$, given by (18). This estimator $(\hat{\Sigma}_1^{MM}, \hat{\Sigma}_2^{MM})$ has constant risk given by

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

This completes the proof. □

We observe from Theorems 6 and 8 that $(\hat{\Sigma}_1^{BU}, \hat{\Sigma}_2^{BU})$ is not minimax and that $(\hat{\Sigma}_1^{MM}, \hat{\Sigma}_2^{MM})$ dominates $(\hat{\Sigma}_1^{BU}, \hat{\Sigma}_2^{BU})$. This implies that in evaluating the risk performance of an alternative estimator for (Σ_1, Σ_2), the estimator to compare with is $(\hat{\Sigma}_1^{MM}, \hat{\Sigma}_2^{MM})$, not $(\hat{\Sigma}_1^{BU}, \hat{\Sigma}_2^{BU})$.

7 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 $(\hat{\Sigma}_1^{MM}, \hat{\Sigma}_2^{MM})$. 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

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

$$\begin{aligned} (\hat{\Sigma}_1^{BU}, \hat{\Sigma}_2^{BU}) &= (S_1/n_1, S_2/n_2) \\ &= (B^{-1}\Psi^{BU}B'^{-1}, B^{-1}\Phi^{BU}B'^{-1}) \end{aligned}$$

where the j'th diagonal element of the diagonal matrix Ψ^{BU}, Φ^{BU} is $(1 - f_j)/n_1$, 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 $c_j, d_j, j = 1, \ldots, p$, the j'th diagonal element of the diagonal matrix Ψ, Φ is $\psi_j = (1 - f_j)/c_j$, $\phi_j = f_j/d_j$ respectively. We define the adjusted usual estimator to be

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

where, for $j = 1, \ldots, p$, the j'th diagonal element of the diagonal matrix Ψ^{AU}, Φ^{AU} is $\psi_j^{AU} = (1-f_j)/(n_1 - p - 1 + 2j)$, $\phi_j^{AU} = f_j/(n_2 + p + 1 - 2j)$ respectively.
We shall show in this subsection that:

1. Under Stein's loss, $\hat{\Sigma}_1^{AU}$ is a minimax estimator of Σ_1.

2. Similarly under Stein's loss, $\hat{\Sigma}_2^{AU}$ is a minimax estimator of Σ_2.

3. Under loss function L, $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$ dominates the constant risk minimax estimator $(\hat{\Sigma}_1^{MM}, \hat{\Sigma}_2^{MM})$.

Proposition 3 *Under Stein's loss, $\hat{\Sigma}_1^{AU}$ is a minimax estimator of Σ_1.*

PROOF. We observe from the proof of Proposition 1 that the conditions for the Wishart identity hold in this case and

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

For convenience of notation, we write $a_i = 1/(n_1 - p - 1 + 2i)$ for $i = 1, \ldots, p$. Then on simplification, we have

$$\begin{aligned} R_S(\hat{\Sigma}_1^{AU}; \Sigma_1) &= E\{\sum_i[(n_1 - p - 1)a_i + 2\sum_{j<i}\frac{f_i(1-f_j)a_j - f_j(1-f_i)a_i}{f_i - f_j} \\ &\quad +2a_i - \log a_i - \log\chi^2_{n_1-i+1} - 1]\} \\ &= E\{\sum_i[(n_1 - p + 1)a_i + 2\sum_{j<i}\frac{f_i(1-f_j)(a_j - a_i)}{f_i - f_j} \\ &\quad +2\sum_{j<i}\frac{[f_i(1-f_j) - f_j(1-f_i)]a_i}{f_i - f_j} \\ &\quad - \log a_i - \log\chi^2_{n_1-i+1} - 1]\}. \end{aligned}$$

Since $f_i(1-f_j)/(f_i - f_j) \geq 1$ whenever $j > i$, we have

$$\begin{aligned} & R_S(\hat{\Sigma}_1^{AU}; \Sigma_1) \\ \leq\ & E\{\sum_i[(n_1 - p + 1)a_i + 2\sum_{j<i}a_i - \log a_i - \log\chi^2_{n_1-i+1} - 1]\} \end{aligned}$$

$$\begin{aligned}
&= E\{\sum_i[(n_1-p-1+2i)a_i - \log a_i - \log\chi^2_{n_1-i+1} - 1]\} \\
&= E\sum_i[(\log(n_1-p-1+2i) - \log\chi^2_{n_1-i+1}] \\
&= E\sum_i[(\log(n_1+p+1-2i) - \log\chi^2_{n_1-i+1}].
\end{aligned}$$

It follows now from Theorem 7 that the right hand side of the last equality is the minimax risk. Hence $\hat{\Sigma}_1^{AU}$ is a minimax estimator of Σ_1. □

Proposition 4 *Under Stein's loss, $\hat{\Sigma}_2^{AU}$ is a minimax estimator of Σ_2.*

PROOF. The proof is similar to that of Proposition 3 and hence will be omitted. □

Theorem 9 *With respect to the loss function L, $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$ dominates $(\hat{\Sigma}_1^{MM}, \hat{\Sigma}_2^{MM})$. Hence $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$ is a minimax estimator of (Σ_1, Σ_2).*

PROOF. This theorem follows directly from Theorem 8 and Propositions 3 and 4 since the loss function under consideration is the sum of the respective loss functions of these two problems. □

7.2 Dey-Srinivasan type Estimators

In the estimation of a covariance matrix, Dey and Srinivasan (1985) constructed a class of minimax estimators for Σ by using a technique of Berger (1980). 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

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

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

$$\begin{aligned}
\psi_i^{DS} &= \frac{1-f_i}{n_1-p-1+2i} + (1-f_i)\beta_i, \\
\phi_i^{DS} &= \frac{f_i}{n_2+p+1-2i} + f_i\gamma_i, \\
u &= \sum_j \log^2(\frac{f_j}{1-f_j}),
\end{aligned}$$

$$\begin{aligned}\beta_i &= [a(u)\log(\frac{f_i}{1-f_i})]/(c+u), \\ \gamma_i &= [b(u)\log(\frac{1-f_i}{f_i})]/(d+u),\end{aligned}$$

$a, b : R^+ \to R$ being suitable functions and c, d being suitable constants. In particular, we shall show that under suitable conditions on a, b, c and d:

1. $\hat{\Sigma}_1^{DS}$ dominates $\hat{\Sigma}_1^{AU}$ in the estimation of Σ_1 under Stein's loss.

2. $\hat{\Sigma}_2^{DS}$ dominates $\hat{\Sigma}_2^{AU}$ in the estimation of Σ_2 under Stein's loss.

3. $(\hat{\Sigma}_1^{DS}, \hat{\Sigma}_2^{DS})$ dominates $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$ in the estimation of (Σ_1, Σ_2) under the loss function L given in (1).

Proposition 5 *In the estimation of Σ_1 under Stein's loss, $\hat{\Sigma}_1^{DS}$ dominates $\hat{\Sigma}_1^{AU}$ whenever*

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

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

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

$$\begin{aligned}|\,(n_1-p-1+2i)\beta_i\,| &= a(u)(n_1-p-1+2i)\,|\log\frac{f_i}{1-f_i}\,|/(c+u) \\ &\leq a(u)(n_1-p-1+2i)/(2\sqrt{c}) \\ &\leq a(u)(n_1+p-1)/(2\sqrt{c}).\end{aligned}$$

Since $\sup_{u\geq 0} a(u)(n_1+p-1)/(2\sqrt{c}) = x < 1$, it follows from Lemma 2.2 of Dey and Srinivasan (1985) that

$$\begin{aligned}-\log[1+(n_1-p-1+2i)\beta_i] \leq\ & (n_1-p-1+2i)^2\beta_i^2(3-x)/[6(1-x)] \\ & -(n_1-p-1+2i)\beta_i. \end{aligned} \tag{19}$$

We observe from the proof of Proposition 1 that the conditions for the Wishart identity are satisfied and furthermore that the difference in risk between $\hat{\Sigma}_1^{DS}$ and $\hat{\Sigma}_1^{AU}$ is given by

$$\begin{aligned} & R_S(\hat{\Sigma}_1^{DS};\Sigma_1) - R_S(\hat{\Sigma}_1^{AU};\Sigma_1) \\ = \; & E\sum_i \{(n_1-p+1)\beta_i + 2\sum_{j<i}\frac{f_j(1-f_i)\beta_i - f_i(1-f_j)\beta_j}{f_j-f_i} \\ & -2f_i(1-f_i)\frac{\partial\beta_i}{\partial f_i} - \log[1+(n_1-p-1+2i)\beta_i]\}. \end{aligned}$$

Now it follows from (19) that

$$\begin{aligned} & R_S(\hat{\Sigma}_1^{DS};\Sigma_1) - R_S(\hat{\Sigma}_1^{AU};\Sigma_1) \\ \leq \; & E\sum_i [(n_1-p+1)\beta_i - 2f_i(1-f_i)\frac{\partial\beta_i}{\partial f_i} - (n_1-p-1+2i)\beta_i \\ & +2\sum_{j<i}\frac{f_j(1-f_i)\beta_i - f_i(1-f_j)\beta_j}{f_j-f_i} + \frac{3-x}{6(1-x)}(n_1-p-1+2i)^2\beta_i^2] \\ = \; & E\sum_i [(2-2i)\beta_i + 2\sum_{j<i}\frac{f_j(1-f_i)\beta_i - f_i(1-f_j)\beta_j}{f_j-f_i} \\ & -2f_i(1-f_i)\frac{\partial\beta_i}{\partial f_i} + \frac{3-x}{6(1-x)}(n_1-p-1+2i)^2\beta_i^2]. \end{aligned} \tag{20}$$

Next we observe that

$$\begin{aligned} & \sum_i [(2-2i)\beta_i + 2\sum_{j<i}\frac{f_j(1-f_i)\beta_i - f_i(1-f_j)\beta_j}{f_j-f_i}] \\ & = \sum_i [(2-2i)\beta_i + 2\sum_{j<i}\frac{f_i(1-f_j)(\beta_i-\beta_j)}{f_j-f_i} + 2\sum_{j<i}\beta_i] \\ & = 2\sum_i\sum_{j<i}\frac{f_i(1-f_j)(\beta_i-\beta_j)}{f_j-f_i} \\ & \leq 0. \end{aligned} \tag{21}$$

Also, we observe that

$$\begin{aligned} & \sum_i [-2f_i(1-f_i)\frac{\partial\beta_i}{\partial f_i} + \frac{3-x}{6(1-x)}(n_1-p-1+2i)^2\beta_i^2] \\ = \; & \sum_i \{-\frac{2f_i(1-f_i)}{(c+u)^2}[(c+u)(2a'(u)\log^2(\frac{f_i}{1-f_i})(\frac{1}{f_i}+\frac{1}{1-f_i}) \end{aligned}$$

$$
\begin{aligned}
& +a(u)(\frac{1}{f_i}+\frac{1}{1-f_i})) - 2a(u)\log^2(\frac{f_i}{1-f_i})(\frac{1}{f_i}+\frac{1}{1-f_i})] \\
& +\frac{3-x}{6(1-x)(c+u)^2}a^2(u)(n_1-p-1+2i)^2\log^2(\frac{f_i}{1-f_i})\} \\
\leq\ & \frac{a(u)}{c+u}[-2p+4+\frac{3-x}{6(1-x)}a(u)(n_1+p-1)^2] \\
& -\frac{4a'(u)}{c+u}\sum_i \log^2(\frac{f_i}{1-f_i}) \\
\leq\ & 0, \qquad (22)
\end{aligned}
$$

since a is nondecreasing and $a(u)(3-x)/[6(1-x)] \leq 2(p-2)/(n_1+p-1)^2$. We now conclude from (20), (21) and (22) that $\hat{\Sigma}_1^{DS}$ dominates $\hat{\Sigma}_1^{AU}$. Minimaxity follows from Theorem 9. □

Corollary 2 *In the estimation of Σ_1 under Stein's loss, $\hat{\Sigma}_1^{DS}$ dominates $\hat{\Sigma}_1^{AU}$ whenever*

1. $0 \leq a(u) \leq 12(p-2)/[5(n_1+p-1)^2]$ *and* $a'(u) \geq 0$ *for all* $u \geq 0$,
2. $c \geq 144(p-2)^2/[25(n_1+p-1)^2]$.

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

PROOF. This follows immediately from Proposition 5. □

Analogous to Proposition 5 and Corollary 2, we have the following two results. Their proofs are very similar to the above-mentioned and are omitted.

Proposition 6 *In the estimation of Σ_2 under Stein's loss, $\hat{\Sigma}_2^{DS}$ dominates $\hat{\Sigma}_2^{AU}$ whenever*

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

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

Corollary 3 *In the estimation of Σ_2 under Stein's loss, $\hat{\Sigma}_2^{DS}$ dominates $\hat{\Sigma}_2^{AU}$ whenever*

1. $0 \leq b(u) \leq 12(p-2)/[5(n_2+p-1)^2]$, *and* $b'(u) \geq 0$ *for all* $u \geq 0$,

2. $d \geq 144(p-2)^2/[25(n_2+p-1)^2]$.

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

Now we shall state the main result of this subsection.

Theorem 10 *In the estimation of* (Σ_1, Σ_2) *under the loss function* L, $(\hat{\Sigma}_1^{DS}, \hat{\Sigma}_2^{DS})$ *dominates* $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$ *whenever*

1. $a(u) \geq 0$ *and* $a'(u) \geq 0$ *for all* $u \geq 0$,
2. $b(u) \geq 0$ *and* $b'(u) \geq 0$ *for all* $u \geq 0$,
3. $\sup_{u \geq 0} a(u)(n_1+p-1)/(2\sqrt{c}) = x < 1$,
4. $\sup_{u \geq 0} b(u)(n_2+p-1)/(2\sqrt{d}) = y < 1$,
5. $\sup_{u \geq 0} a(u)(3-x)/[6(1-x)] \leq 2(p-2)/(n_1+p-1)^2$,
6. $\sup_{u \geq 0} b(u)(3-y)/[6(1-y)] \leq 2(p-2)/(n_2+p-1)^2$.

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

PROOF. This theorem follows directly from Propositions 5 and 6 since the loss function under consideration is the sum of the respective loss functions of these two problems. Minimaxity follows immediately from Theorem 9. □

Corollary 4 *In the estimation of* (Σ_1, Σ_2) *under the loss function given by (1),* $(\hat{\Sigma}_1^{DS}, \hat{\Sigma}_2^{DS})$ *dominates* $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$ *whenever*

1. $0 \leq a(u) \leq 12(p-2)/[5(n_1+p-1)^2]$, *and* $a'(u) \geq 0$ *for all* $u \geq 0$,
2. $0 \leq b(u) \leq 12(p-2)/[5(n_2+p-1)^2]$, *and* $b'(u) \geq 0$ *for all* $u \geq 0$,
3. $c \geq 144(p-2)^2/[25(n_1+p-1)^2]$, $\quad d \geq 144(p-2)^2/[25(n_2+p-1)^2]$.

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

PROOF. This follows immediately from Theorem 10. □

Analogous to Dey and Srinivasan (1985), one can construct adapted versions of these minimax estimators. We let

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

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

$$\begin{aligned}
\psi_i^{AD} &= \frac{1-f_i}{n_1-p-1+2i} + (1-f_i)\beta_i, \\
\phi_i^{AD} &= \frac{f_i}{n_2+p+1-2i} + f_i\gamma_i, \\
u &= \sum_j [\log(\frac{f_j}{1-f_j}) - \frac{1}{p}\sum_k \log(\frac{f_k}{1-f_k})]^2, \\
\beta_i &= a(u)[\log(\frac{f_i}{1-f_i}) - \frac{1}{p}\sum_k \log(\frac{f_k}{1-f_k})]/(c+u), \\
\gamma_i &= b(u)[\log(\frac{1-f_i}{f_i}) - \frac{1}{p}\sum_k \log(\frac{f_k}{1-f_k})]/(d+u),
\end{aligned}$$

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

We shall now state the results for these estimators. The proofs are very similar to those of Propositions 5, 6 and Theorem 10 and hence are omitted.

Proposition 7 *In the estimation of* Σ_1 *under Stein's loss,* $\hat{\Sigma}_1^{AD}$ *dominates* $\hat{\Sigma}_1^{AU}$ *whenever*

1. $a(u) \geq 0$ *and* $a'(u) \geq 0$ *for all* $u \geq 0$,
2. $\sup_{u\geq 0} a(u)(n_1+p-1)/(2\sqrt{c}) = x < 1$,
3. $\sup_{u\geq 0} a(u)(3-x)/[6(1-x)] \leq [2(p-3)+4/p]/(n_1+p-1)^2$.

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

Proposition 8 *In the estimation of* Σ_2 *under Stein's loss,* $\hat{\Sigma}_2^{AD}$ *dominates* $\hat{\Sigma}_2^{AU}$ *whenever*

1. $b(u) \geq 0$ *and* $b'(u) \geq 0$ *for all* $u \geq 0$,
2. $\sup_{u\geq 0} b(u)(n_2+p-1)/(2\sqrt{d}) = y < 1$,
3. $\sup_{u\geq 0} b(u)(3-y)/[6(1-y)] \leq [2(p-3)+4/p]/(n_2+p-1)^2$.

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

Theorem 11 *In the estimation of* (Σ_1, Σ_2) *under the loss function* L, $(\hat{\Sigma}_1^{AD}, \hat{\Sigma}_2^{AD})$ *dominates* $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$ *whenever*

1. $a(u) \geq 0$ *and* $a'(u) \geq 0$ *for all* $u \geq 0$,

2. $b(u) \geq 0$ *and* $b'(u) \geq 0$ *for all* $u \geq 0$,
3. $\sup_{u\geq 0} a(u)(n_1 + p - 1)/(2\sqrt{c}) = x < 1$,
4. $\sup_{u\geq 0} b(u)(n_2 + p - 1)/(2\sqrt{d}) = y < 1$,
5. $\sup_{u\geq 0} a(u)(3 - x)/[6(1 - x)] \leq [2(p - 2) + 4/p]/(n_1 + p - 1)^2$,
6. $\sup_{u\geq 0} b(u)(3 - y)/[6(1 - y)] \leq [2(p - 2) + 4/p]/(n_2 + p - 1)^2$.

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

7.3 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 $(\hat{\Sigma}_1^{ST}, \hat{\Sigma}_2^{ST})$ for (Σ_1, Σ_2). Let $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ be an estimator for (Σ_1, Σ_2) where

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

$\Phi = \text{diag}(\phi_1, \ldots, \phi_p)$, $\Psi = \text{diag}(\psi_1, \ldots, \psi_p)$, $B(S_1+S_2)B' = I$, $BS_2B' = F = \text{diag}(f_1, \ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$. Under loss function L, we observe from Theorem 5 that

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

$$-\log \chi^2_{n_1-i+1} + \frac{n_2-p+1}{f_i}\phi_i + 2\phi_i \sum_{j\neq i} \frac{1-f_j}{f_i-f_j}$$
$$+2f_i(1-f_i)\frac{\partial}{\partial f_i}(\frac{\phi_i}{f_i}) - \log\frac{\phi_i}{f_i} - \log \chi^2_{n_2-i+1} - 2]\}.$$

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

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

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

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

On simplification, we have

$$\begin{aligned}\psi_i &= (1-f_i)/[n_1-p+1-2\sum_{j\neq i}\frac{f_j(1-f_i)}{f_i-f_j}], \\ \phi_i &= f_i/[n_2-p+1+2\sum_{j\neq i}\frac{f_i(1-f_j)}{f_i-f_j}], \qquad i=1,\ldots,p.\end{aligned} \tag{23}$$

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

$$0 \le \psi_1 \le \cdots \le \psi_p,$$
$$\phi_1 \ge \cdots \ge \phi_p \ge 0.$$

However with the ψ_i's and ϕ_i's defined by (23), 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}, $i = 1, \ldots, p$, which satisfy

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

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

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

where $\Phi^{ST} = \mathrm{diag}(\phi_1^{ST}, \ldots, \phi_p^{ST})$, $\Psi^{ST} = \mathrm{diag}(\psi_1^{ST}, \ldots, \psi_p^{ST})$, $B(S_1+S_2)B' = I, BS_2B' = F = \mathrm{diag}(f_1, \ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$. This concludes our construction of the Stein-type estimator.

7.4 Haff-type Estimator

Haff (1982), (1988) constructed an estimator for a covariance matrix which has a similar functional form to that of the Stein (1975) estimator. His technique is briefly summarised as follows:

First, a prior distribution is put on the parameter space. Then the average risk of an almost arbitrary orthogonally invariant estimator for Σ is minimized via the Euler-Lagranges equations and the Haff estimator is determined from these equations.

In this subsection, we shall apply Haff's method to obtain an alternative estimator, denoted by $(\hat{\Sigma}_1^{HF}, \hat{\Sigma}_2^{HF})$, for (Σ_1, Σ_2). We note that an equvariant estimator $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ for (Σ_1, Σ_2) must be of the form

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

where $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_p)$, $\Psi = \mathrm{diag}(\psi_1, \ldots, \psi_p)$, $B(S_1+S_2)B' = I, BS_2B' = F = \mathrm{diag}(f_1, \ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$. Then from Theorem 5, under the loss function given by (1), the unbiased estimate of the risk of an almost arbitrary equivariant estimator for (Σ_1, Σ_2) is

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

Since the ψ_i's and ϕ_i's follow the ordering:

$$\begin{aligned}
&0 \leq \psi_1 \leq \cdots \leq \psi_p, \\
&\phi_1 \geq \cdots \geq \phi_p \geq 0,
\end{aligned}$$

we write for each i,

$$\begin{aligned}\psi_i(F) &= \sum_{k\geq i}\varepsilon_k^2(F),\\ \phi_i(F) &= \sum_{k\geq i}\epsilon_k^2(F).\end{aligned}$$

This implies that $\hat{R}$ can be expressed as

$$\begin{aligned}\hat{R} = & \sum_i\Big[\frac{n_1-p-1}{1-f_i}\sum_{k\geq i}\varepsilon_k^2 - 2\sum_{k\geq i}\varepsilon_k^2\sum_{j\neq i}\frac{f_j}{f_i-f_j}\\ & +2\sum_{k\geq i}\varepsilon_k^2 + 2f_i\sum_{k\geq i}\frac{\partial\varepsilon_k^2}{\partial(1-f_i)} - \log\Big(\sum_{k\geq i}\frac{\varepsilon_k^2}{1-f_i}\Big)\\ & +\frac{n_2-p-1}{f_i}\sum_{k\geq i}\epsilon_k^2 + 2\sum_{k\geq i}\epsilon_k^2\sum_{j\neq i}\frac{1-f_j}{f_i-f_j}\\ & +2\sum_{k\geq i}\epsilon_k^2 + 2(1-f_i)\sum_{k\geq i}\frac{\partial\epsilon_k^2}{\partial f_i} - \log\Big(\sum_{k\geq i}\frac{\epsilon_k^2}{f_i}\Big)\\ & -E\log\chi^2_{n_1-i+1} - E\log\chi^2_{n_2-i+1} - 2\Big].\end{aligned}$$

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

$$\int \hat{G}(f_1,\ldots,f_p;\psi_1,\ldots,\psi_p;\phi_1,\ldots,\phi_p;\frac{\partial\psi_1}{\partial f_1},\ldots,\frac{\partial\psi_p}{\partial f_p};\frac{\partial\phi_1}{\partial f_1},\ldots,\frac{\partial\phi_p}{\partial f_p})\mathrm{d}F,$$

where

$$\hat{G} = m\hat{R}.$$

The solution of the Euler-Lagranges equations minimizes the average risk. These equations are

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

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

$$(24)\quad \varepsilon_k\sum_{i\leq k}\Big[\frac{n_1-p-1}{1-f_i} - 2\sum_{j\neq i}\frac{f_j}{f_i-f_j} + 4 - \psi_i^{-1} + 2f_i\frac{\partial\log m}{\partial f_i}\Big] = 0,$$

and

$$\epsilon_k \sum_{i \le k} [\frac{n_2 - p - 1}{f_i} + 2 \sum_{j \ne i} \frac{1 - f_j}{f_i - f_j} + 4 - \phi_i^{-1} - 2(1 - f_i) \frac{\partial \log m}{\partial f_i}] = 0.$$

(25)

Next we set

(26) $$m(F) = \prod_i 1/[f_i(1 - f_i)].$$

This is motivated by the observation that in order for the estimator ($\hat{\Sigma}_1^{HF}$, $\hat{\Sigma}_2^{HF}$) to compete favorably with ($\hat{\Sigma}_1^{MM}, \hat{\Sigma}_2^{MM}$) and ($\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU}$), the form of ($\hat{\Sigma}_1^{HF}$, $\hat{\Sigma}_2^{HF}$) should approach that of ($\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU}$) when the eigenvalues of $\Sigma_2(\Sigma_1 + \Sigma_2)^{-1}$ are far apart. Substituting (26) into (24) and (25), we get

$$\varepsilon_k \sum_{i \le k} [\frac{n_1 - p + 1}{1 - f_i} - 2 \sum_{j \ne i} \frac{f_j}{f_i - f_j} - \psi_i^{-1}] = 0,$$

and

$$\epsilon_k \sum_{i \le k} [\frac{n_2 - p + 1}{f_i} + 2 \sum_{j \ne i} \frac{1 - f_j}{f_i - f_j} - \phi_i^{-1}] = 0.$$

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 ε_i^{HF}, ϵ_i^{HF}, $1 \le i \le p$ and write

$$\begin{aligned} \psi_i^{HF}(F) &= \sum_{k \ge i} \varepsilon_k^{HF^2}(F), \\ \phi_i^{HF}(F) &= \sum_{k \ge i} \epsilon_k^{HF^2}(F). \end{aligned}$$

It is clear that the natural ordering of the ψ_i^{HF}'s and ϕ_i^{HF}'s is satisfied in this case. That is

$$\begin{gathered} 0 \le \psi_1^{HF} \le \cdots \le \psi_p^{HF}, \\ \phi_1^{HF} \ge \cdots \ge \phi_p^{HF} \ge 0. \end{gathered}$$

We now define

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

where $\Phi^{HF} = \text{diag}(\phi_1^{HF}, \ldots, \phi_p^{HF})$, $\Psi^{HF} = \text{diag}(\psi_1^{HF}, \ldots, \psi_p^{HF})$, $B(S_1 + S_2)B' = I, BS_2B' = F = \text{diag}(f_1, \ldots, f_p)$ with $f_1 \ge \cdots \ge f_p$. This completes our construction of the Haff-type estimator for (Σ_1, Σ_2).

8 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. In Tables 1 to 4, 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 $(\hat{\Sigma}_1^{BU}, \hat{\Sigma}_2^{BU}) =$ BU, $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU}) =$ AU, etc.

As it is, the estimator $(\hat{\Sigma}_1^{DS}, \hat{\Sigma}_2^{DS})$ is not well-defined. In this study we take

$$\begin{aligned} a &= 6(p-2)/[5(n_1+p-1)^2], \\ b &= 6(p-2)/[5(n_2+p-1)^2], \\ c &= 5.8(p-2)^2/(n_1+p-1)^2, \\ d &= 5.8(p-2)^2/(n_2+p-1)^2. \end{aligned}$$

These values are chosen with the aim of doing well when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are close together.

In Tables 5 to 8, under Stein's loss, the average losses and their estimated standard deviations based on 500 independent replications are calculated for the following estimators of Σ_i: $\hat{\Sigma}_i^{MM}$, $\hat{\Sigma}_i^{ST}$ and $\hat{\Sigma}_i^{HF}$.

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 to one.

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. 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. It has been proved that $(\hat{\Sigma}_1^{DS}, \hat{\Sigma}_2^{DS})$ dominates $(\hat{\Sigma}_1^{AU}, \hat{\Sigma}_2^{AU})$. However this Monte Carlo study reveals that the difference in risk between these two estimators is rather small at best.

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.

We also 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 conservative indicator 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 and comments. Finally I would also like to thank Professors Dipak Dey and L. R. Haff for sending me copies of their unpublished manuscripts.

TABLE 1

$n_1 = 12 \qquad n_2 = 12$

Average losses of estimators for the estimation of (Σ_1, Σ_2)

(Estimated standard errors are in parenthesis)

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)

TABLE 2

$n_1 = 25 \qquad n_2 = 25$

Average losses of estimators for the estimation of (Σ_1, Σ_2)

(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	BU	MM	AU	DS	ST	HF
(1,1,1,1,1,	5.18	4.62	3.47	3.39	2.82	2.80
1,1,1,1,1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
(10,10,10,10,10,	5.18	4.62	3.98	3.96	3.80	3.80
1,1,1,1,1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(25,25,25,25,25,	5.18	4.62	3.86	3.85	3.91	3.91
25,25,25,1,1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(30,30,30,1,1,	5.18	4.62	3.99	3.98	3.69	3.68
1,1,1,1,1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(50,1,1,1,1,	5.18	4.62	3.71	3.68	3.21	3.19
1,1,1,1,1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(20,20,20,5,5,	5.18	4.62	4.03	4.01	3.96	3.96
5,5,1,1,1)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
(512,256,128,64,32,	5.18	4.62	4.31	4.30	4.40	4.40
16,8,4,2,1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(0.50,0.45,0.40,0.35,0.30,	5.18	4.62	3.75	3.73	3.55	3.55
0.25,0.20,0.15,0.10,0.05)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
(10,0.1,0.1,0.1,0.1,	5.18	4.62	3.71	3.70	3.68	3.67
0.1,0.1,0.1,0.1,0.1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(10,5,1,0.1,0.1,	5.18	4.62	4.03	4.02	3.97	3.97
0.1,0.1,0.1,0.1,0.1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
(10,10,10,0.1,0.1,	5.18	4.62	4.02	4.01	3.90	3.90
0.1,0.1,0.1,0.1,0.1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(10,10,10,10,1,	5.18	4.62	4.23	4.22	4.17	4.17
1,1,0.1,0.1,0.1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(10,10,5,5,1,	5.18	4.62	4.21	4.20	4.20	4.21
1,0.4,0.4,0.1,0.1)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(20,6,8/3,14/9,1,	5.18	4.62	4.19	4.18	4.21	4.21
2/3,4/9,2/7,1/6,2/27)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(81,27,9,3,1,	5.18	4.62	4.39	4.39	4.46	4.47
1/2,1/4,1/8,1/16,1/32)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(10^3,10^2,25,5,2,	5.18	4.62	4.54	4.54	4.60	4.60
1/2,1/5,1/20,10^{-2},10^{-3})	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

TABLE 3

$n_1 = 12 \qquad n_2 = 25$

Average losses of estimators for the estimation of (Σ_1, Σ_2)

(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	BU	MM	AU	DS	ST	HF
(1,1,1,1,1,	9.88	8.09	5.91	5.87	4.89	4.72
1,1,1,1,1)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
(10,10,10,10,10,	9.88	8.09	6.59	6.57	6.19	6.19
1,1,1,1,1)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
(25,25,25,25,25,	9.88	8.09	6.40	6.40	5.98	5.92
25,25,25,1,1)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
(30,30,30,1,1,	9.88	8.09	6.85	6.83	6.57	6.58
1,1,1,1,1)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
(50,1,1,1,1,	9.88	8.09	6.56	6.54	6.13	6.09
1,1,1,1,1)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
(20,20,20,5,5,	9.88	8.09	6.62	6.61	6.37	6.38
5,5,1,1,1)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
(512,256,128,64,32,	9.88	8.09	7.15	7.15	7.26	7.29
16,8,4,2,1)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(0.50,0.45,0.40,0.35,0.30,	9.88	8.09	6.22	6.18	5.68	5.56
0.25,0.20,0.15,0.10,0.05)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
(10,0.1,0.1,0.1,0.1,	9.88	8.09	6.61	6.59	6.77	6.73
0.1,0.1,0.1,0.1,0.1)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
(10,5,1,0.1,0.1,	9.88	8.09	6.99	6.97	7.03	7.05
0.1,0.1,0.1,0.1,0.1)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(10,10,10,0.1,0.1,	9.88	8.09	6.97	6.95	6.85	6.86
0.1,0.1,0.1,0.1,0.1)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
(10,10,10,10,1,	9.88	8.09	7.02	7.00	6.85	6.86
1,1,0.1,0.1,0.1)	(0.07)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)
(10,10,5,5,1,	9.88	8.09	6.94	6.93	6.79	6.80
1,0.4,0.4,0.1,0.1)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(20,6,8/3,14/9,1,	9.88	8.09	7.00	7.00	6.95	6.97
2/3,4/9,2/7,1/6,2/27)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(81,27,9,3,1,	9.88	8.09	7.42	7.41	7.50	7.53
1/2,1/4,1/8,1/16,1/32)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(10^3,10^2,25,5,2,	9.88	8.09	7.78	7.78	7.93	7.94
1/2,1/5,1/20,10^{-2},10^{-3})	(0.07)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)

TABLE 4

$n_1 = 25 \qquad n_2 = 12$

Average losses of estimators for the estimation of (Σ_1, Σ_2)

(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	BU	MM	AU	DS	ST	HF
(1,1,1,1,1,	9.88	8.06	5.93	5.89	4.92	4.76
1,1,1,1,1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(10,10,10,10,10,	9.88	8.06	6.60	6.57	6.18	6.11
1,1,1,1,1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(25,25,25,25,25,	9.88	8.06	6.67	6.65	6.99	6.98
25,25,25,1,1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(30,30,30,1,1,	9.88	8.06	6.61	6.59	6.05	5.95
1,1,1,1,1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(50,1,1,1,1,	9.88	8.06	6.22	6.19	5.37	5.23
1,1,1,1,1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(20,20,20,5,5,	9.88	8.06	6.64	6.61	6.32	6.27
5,5,1,1,1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(512,256,128,64,32,	9.88	8.06	7.14	7.13	7.17	7.15
16,8,4,2,1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(0.50,0.45,0.40,0.35,0.30,	9.88	8.06	6.27	6.25	5.84	5.82
0.25,0.20,0.15,0.10,0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(10,0.1,0.1,0.1,0.1,	9.88	8.06	6.22	6.21	5.73	5.66
0.1,0.1,0.1,0.1,0.1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(10,5,1,0.1,0.1,	9.88	8.06	6.63	6.62	6.21	6.15
0.1,0.1,0.1,0.1,0.1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(10,10,10,0.1,0.1,	9.88	8.06	6.66	6.65	6.20	6.14
0.1,0.1,0.1,0.1,0.1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(10,10,10,10,1,	9.88	8.06	7.08	7.07	6.96	6.96
1,1,0.1,0.1,0.1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(10,10,5,5,1,	9.88	8.06	6.97	6.95	6.84	6.84
1,0.4,0.4,0.1,0.1)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(20,6,8/3,14/9,1,	9.88	8.06	6.89	6.88	6.74	6.75
2/3,4/9,2/7,1/6,2/27)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(81,27,9,3,1,	9.88	8.06	7.27	7.26	7.26	7.27
1/2,1/4,1/8,1/16,1/32)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(10^3,10^2,25,5,2,	9.88	8.06	7.74	7.73	7.86	7.87
1/2,1/5,1/20,10^{-2},10^{-3})	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)

TABLE 5

$n_1 = 12 \qquad n_2 = 12$

Average losses of estimators for the estimation of Σ_i, i=1, 2

(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	Estimating Σ_1			Estimating Σ_2		
	$\hat{\Sigma}_1^{MM}$	$\hat{\Sigma}_1^{ST}$	$\hat{\Sigma}_1^{HF}$	$\hat{\Sigma}_2^{MM}$	$\hat{\Sigma}_2^{ST}$	$\hat{\Sigma}_2^{HF}$
(25,25,25,25,25,	5.83	4.27	4.26	5.79	5.45	5.44
25,25,25,1,1)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
(100,90,80,70,60,	5.83	4.23	4.23	5.79	6.31	6.28
50,40,30,20,10)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
(512,256,128,64,32,	5.83	5.08	5.12	5.79	5.34	5.32
16,8,4,2,1)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
(10,10,10,5,1,	5.83	4.99	4.98	5.79	4.79	4.76
0.1,0.1,0.1,0.1,0.1)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
(10,10,10,10,1,	5.83	4.85	4.85	5.79	5.03	5.03
1,1,0.1,0.1,0.1)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
(10,10,5,5,1,	5.83	4.83	4.83	5.79	4.90	4.89
1,0.4,0.4,0.1,0.1)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
(20,6,8/3,14/9,1,	5.83	4.98	5.00	5.79	4.77	4.77
2/3,4/9,2/7,1/6,2/27)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
(81,27,9,3,1,	5.83	5.34	5.36	5.79	5.11	5.12
1/2,1/4,1/8,1/16,1/32)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
(10^3,10^2,25,5,2,	5.83	5.62	5.63	5.79	5.61	5.62
1/2,1/5,1/20,10^{-2},10^{-3})	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)

TABLE 6

$n_1 = 25 \qquad n_2 = 25$

Average losses of estimators for the estimation of Σ_i, i=1, 2

(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	Estimating Σ_1 $\hat{\Sigma}_1^{MM}$	$\hat{\Sigma}_1^{ST}$	$\hat{\Sigma}_1^{HF}$	Estimating Σ_2 $\hat{\Sigma}_2^{MM}$	$\hat{\Sigma}_2^{ST}$	$\hat{\Sigma}_2^{HF}$
(25,25,25,25,25,	2.30	1.68	1.67	2.32	2.24	2.24
25,25,25,1,1)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
(100,90,80,70,60,	2.30	1.71	1.71	2.32	2.41	2.40
50,40,30,20,10)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
(512,256,128,64,32,	2.30	2.20	2.20	2.32	2.20	2.20
16,8,4,2,1)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
(10,10,10,5,1,	2.30	2.10	2.10	2.32	2.01	2.01
0.1,0.1,0.1,0.1,0.1)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
(10,10,10,10,1,	2.30	2.05	2.05	2.32	2.12	2.12
1,1,0.1,0.1,0.1)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
(10,10,5,5,1,	2.30	2.08	2.08	2.32	2.13	2.13
1,0.4,0.4,0.1,0.1)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
(20,6,8/3,14/9,1,	2.30	2.11	2.11	2.32	2.10	2.10
2/3,4/9,2/7,1/6,2/27)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
(81,27,9,3,1,	2.30	2.24	2.25	2.32	2.22	2.22
1/2,1/4,1/8,1/16,1/32)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
(10^3,10^2,25,5,2,	2.30	2.29	2.29	2.32	2.31	2.31
1/2,1/5,1/20,10^{-2},10^{-3})	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)

TABLE 7

$n_1 = 12 \qquad n_2 = 25$

Average losses of estimators for the estimation of Σ_i, i=1, 2

(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	Estimating Σ_1 $\hat{\Sigma}_1^{MM}$	$\hat{\Sigma}_1^{ST}$	$\hat{\Sigma}_1^{HF}$	Estimating Σ_2 $\hat{\Sigma}_2^{MM}$	$\hat{\Sigma}_2^{ST}$	$\hat{\Sigma}_2^{HF}$
(25,25,25,25,25,	5.79	3.68	3.62	2.30	2.30	2.30
25,25,25,1,1)	(0.06)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
(100,90,80,70,60,	5.79	3.63	3.59	2.30	2.48	2.48
50,40,30,20,10)	(0.06)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
(512,256,128,64,32,	5.79	4.99	5.02	2.30	2.27	2.27
16,8,4,2,1)	(0.06)	(0.05)	(0.05)	(0.05)	(0.02)	(0.02)
(10,10,10,5,1,	5.79	4.78	4.78	2.30	2.12	2.12
0.1,0.1,0.1,0.1,0.1)	(0.06)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
(10,10,10,10,1,	5.79	4.68	4.69	2.30	2.18	2.18
1,1,0.1,0.1,0.1)	(0.06)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
(10,10,5,5,1,	5.79	4.64	4.65	2.30	2.16	2.16
1,0.4,0.4,0.1,0.1)	(0.06)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
(20,6,8/3,14/9,1,	5.79	4.82	4.85	2.30	2.13	2.13
2/3,4/9,2/7,1/6,2/27)	(0.06)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
(81,27,9,3,1,	5.79	5.29	5.31	2.30	2.21	2.22
1/2,1/4,1/8,1/16,1/32)	(0.06)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
(10^3,10^2,25,5,2,	5.79	5.65	5.66	2.30	2.28	2.28
1/2,1/5,1/20,10^{-2},10^{-3})	(0.06)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)

TABLE 8

$n_1 = 25 \qquad n_2 = 12$

Average losses of estimators for the estimation of Σ_i, i=1, 2

(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	Estimating Σ_1 $\hat{\Sigma}_1^{MM}$	$\hat{\Sigma}_1^{ST}$	$\hat{\Sigma}_1^{HF}$	Estimating Σ_2 $\hat{\Sigma}_2^{MM}$	$\hat{\Sigma}_2^{ST}$	$\hat{\Sigma}_2^{HF}$
(25,25,25,25,25,	2.32	1.98	1.98	5.74	5.02	5.00
25,25,25,1,1)	(0.02)	(0.02)	(0.02)	(0.05)	(0.04)	(0.04)
(100,90,80,70,60,	2.32	1.98	1.98	5.74	5.58	5.52
50,40,30,20,10)	(0.02)	(0.02)	(0.02)	(0.05)	(0.04)	(0.04)
(512,256,128,64,32,	2.32	2.25	2.25	5.74	4.92	4.89
16,8,4,2,1)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.05)
(10,10,10,5,1,	2.32	2.21	2.21	5.74	4.45	4.41
0.1,0.1,0.1,0.1,0.1)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.05)
(10,10,10,10,1,	2.32	2.17	2.17	5.74	4.79	4.79
1,1,0.1,0.1,0.1)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.05)
(10,10,5,5,1,	2.32	2.16	2.16	5.74	4.67	4.68
1,0.4,0.4,0.1,0.1)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.05)
(20,6,8/3,14/9,1,	2.32	2.19	2.19	5.74	4.55	4.56
2/3,4/9,2/7,1/6,2/27)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.05)
(81,27,9,3,1,	2.32	2.28	2.28	5.74	4.97	4.99
1/2,1/4,1/8,1/16,1/32)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.05)
(10^3,10^2,25,5,2,	2.32	2.32	2.32	5.74	5.54	5.55
1/2,1/5,1/20,10^{-2},10^{-3})	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.05)

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] BILODEAU, M. (1987). Empirical Bayes on Wishart-matrix observations. Unpublished manuscript.

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

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

[5] DEY, D. K. and SRINIVASAN, C. (1986). Trimmed minimax estimator of a covariance matrix. *Ann. Inst. Statist. Math.* **38** *A* 47-54.

[6] DIACONIS, P. and STEIN, C. (1983). *Lectures on Statistical Decision Theory.* Unpublished Lecture Notes, Stanford University, Stanford.

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

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

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

[10] 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.

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

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

[13] HAFF, L. R. (1981). Further identities for the Wishart distribution with applications in regression. *Canadian J. of Statist.* **9** 215-224.

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

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

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

[17] 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.

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

[19] OLKIN, I. and SELLIAH, J. B. (1977). Estimating covariances in a multivariate normal distribution. *Statistical Decision Theory and Related Topics II*, 313-326, ed. S. S. Gupta & D. Moore. Academic Press, New York.

[20] STEIN, C. (1956). Some problems in multivariate analysis. Part I. Technical Report No. 6, Department of Statistics, Stanford University.

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

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

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

[24] 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.

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