The Annals of Statistics
1991, Vol. 19, No. 1, 297–313

ESTIMATING THE COMMON MEAN OF TWO MULTIVARIATE NORMAL DISTRIBUTIONS

By Wei-Liem Loh

Purdue University

Let X_1, X_2 be two $p \times 1$ multivariate normal random vectors and S_1, S_2 be two $p \times p$ Wishart matrices, where $X_1 \sim N_p(\xi, \Sigma_1)$, $X_2 \sim N_p(\xi, \Sigma_2)$, $S_1 \sim W_p(\Sigma_1, n)$ and $S_2 \sim W_p(\Sigma_2, n)$. We further assume that X_1, X_2, S_1, S_2 are stochastically independent. We wish to estimate the common mean ξ with respect to the loss function $L = (\hat{\xi} - \xi)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \xi)$. By extending the methods of Stein and Haff, an alternative unbiased estimator to the usual generalized least squares estimator is obtained. However, the risk of this estimator is not available in closed form. A Monte Carlo swindle is used instead to evaluate its risk performance. The results indicate that the alternative estimator performs very favorably against the usual estimator.

1. Introduction. In this paper we consider the problem of estimating the common mean of two multivariate normal distributions with unknown covariance matrices. The precise formulation of this problem is as follows:

Let X_1, X_2 be two $p \times 1$ multivariate normal random vectors and S_1, S_2 be two $p \times p$ Wishart matrices, where $X_1 \sim N_p(\xi, \Sigma_1)$, $X_2 \sim N_p(\xi, \Sigma_2)$, $S_1 \sim W_p(\Sigma_1, n)$, $S_2 \sim W_p(\Sigma_2, n)$ with X_1, X_2, S_1, S_2 mutually independent and ξ, Σ_1, Σ_2 unknown. We wish to estimate ξ under the quadratic loss function:

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

The above loss function is a natural symmetric extension of the following invariant loss function:

$$L_0(\hat{\xi}; \xi, \Sigma) = (\hat{\xi} - \xi)'\Sigma^{-1}(\hat{\xi} - \xi),$$

which was first considered by James and Stein (1961) in estimating the mean ξ of a multivariate normal distribution with unknown covariance matrix Σ.

When $p = 1$, there is a lot of research on this problem. In particular, Graybill and Deal (1959) have shown that the unbiased estimator for ξ given by

$$\hat{\xi} = (S_2 X_1 + S_1 X_2)/(S_1 + S_2)$$

has smaller variance than either of X_1 or X_2 if $n > 10$. Other related

Received October 1988; revised September 1989.

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

Key words and phrases. Common mean, equivariant estimation, unbiased estimate of risk, Wishart distribution.

literature includes Brown and Cohen (1974), Cohen and Sackrowitz (1974) and Khatri and Shah (1974).

When $p > 1$, Chiou and Cohen (1985) discuss this problem by evaluating unbiased estimators for ξ using their covariance matrices as a criterion. Also, Shinozaki (1978) considers the estimation of the common mean of k multivariate normal distributions, where the covariance matrices are known up to an arbitrary constant.

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

2. Equivariant estimators.

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

$$\xi \to A\xi + \alpha, \qquad X_i \to AX_i + \alpha,$$

$$\Sigma_i \to A\Sigma_i A', \qquad S_i \to AS_i A',$$

where $\alpha \in R^p$, $A \in GL(p, R)$, the group of $p \times p$ nonsingular matrices and $i = 1, 2$. For simplicity of notation, if $x = (x_1, \dots, x_p)'$ we define $|x|^i$ to be $(|x_1|^i, \dots, |x_p|^i)'$.

THEOREM 1. *Let $X_i \sim N_p(\xi, \Sigma_i)$, $S_i \sim W_p(\Sigma_i, n)$, $i = 1, 2$, with X_1, X_2, S_1, S_2 independent. Then under the group of affine transformations, $\hat{\xi}$ is an equivariant estimator for ξ if and only if $\hat{\xi}$ can be expressed as*

$$\hat{\xi}(X_1, X_2, S_1, S_2) = B^{-1}\Phi B X_1 + B^{-1}(I - \Phi) B X_2, \tag{1}$$

where $\Phi = \Phi(|B(X_1 - X_2)|^2, F)$ is a diagonal matrix, $B(S_1 + S_2)B' = I$, $BS_2B' = F = \operatorname{diag}(f_1, \dots, f_p)$ with $f_1 \geq \cdots \geq f_p$.

PROOF. The proof is straightforward and we refer the reader to Loh (1988) for details. □

The estimators that we shall be considering are special cases of (1).

2.1. *Generalized least squares estimator.* First suppose that the two covariance matrices Σ_1, Σ_2 are known. Then with respect to the loss function

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

the best linear unbiased estimator $\hat{\xi}^{\mathrm{BE}}$ for ξ is given by

$$\hat{\xi}^{\mathrm{BE}} = \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)^{-1}\left(\Sigma_1^{-1}X_1 + \Sigma_2^{-1}X_2\right).$$

The following proposition, which gives some support to the claim that the loss function L is a natural symmetric extension of L_0, is needed in the sequel.

PROPOSITION 1. *With respect to the loss function L, the risk of $\hat{\xi}^{\mathrm{BE}}$ is p.*

PROOF. The proof is straightforward and is omitted. □

However, for the problem that we are concerned with in this paper, the covariance matrices Σ_1, Σ_2 are unknown. The usual practice would be to replace Σ_1, Σ_2 in $\hat{\xi}^{\mathrm{BE}}$ by their maximum likelihood estimators $S_1/n, S_2/n$. This results in the usual estimator $\hat{\xi}^{\mathrm{LS}}$ for ξ which is given by

$$\hat{\xi}^{\mathrm{LS}} = \left(S_1^{-1} + S_2^{-1}\right)^{-1}\left(S_1^{-1}X_1 + S_2^{-1}X_2\right).$$

We observe that $\hat{\xi}^{\mathrm{LS}}$ can be expressed as in (1) with $\Phi = F$. Furthermore, we note that $\hat{\xi}^{\mathrm{LS}}$ is the generalized least squares estimator in the sense that it attains

$$\min_{\hat{\xi}} \sum_i (X_i - \hat{\xi})' S_i^{-1} (X_i - \hat{\xi}).$$

It is well known that the eigenvalues of $(S_1^{-1} + S_2^{-1})^{-1}S_1^{-1}$ are more spread out than the eigenvalues of its expectation. The next estimator for ξ tries to exploit this by correcting this eigenvalue distortion.

2.2. *Stein–Haff type estimator.* With the notation of Theorem 1, we define for $i = 1, \dots, p$,

$$\phi_i = \left[\beta_i^{\mathrm{SH}}/(1 - f_i)\right] / \left\{\left[\alpha_i^{\mathrm{SH}}/f_i\right] + \left[\beta_i^{\mathrm{SH}}/(1 - f_i)\right]\right\}, \tag{2}$$

where

$$\alpha_i^{\mathrm{SH}} = n - p - 1 + 2(1 - f_i) + 2\sum_{j \neq i} \frac{f_i(1 - f_j)}{f_i - f_j},$$

$$\beta_i^{\mathrm{SH}} = n - p - 1 + 2f_i - 2\sum_{j \neq i} \frac{(1 - f_i)f_j}{f_i - f_j}.$$

[The derivation of (2), which is elaborated in Section 5, uses the unbiased estimate of risk techniques of Stein (1975, 1977a, b) and Haff (1982, 1988).] Unfortunately, the natural ordering of the ϕ_i's may be altered. The natural ordering is given by $\phi_1 \geq \cdots \geq \phi_p \geq 0$. To correct for this, Stein's (1975) isotonic regression is applied to the $\alpha_i^{\mathrm{SH}}/f_i$'s and the $\beta_i^{\mathrm{SH}}/(1 - f_i)$'s. This results in φ_i^{SH} and ψ_i^{SH}, $i = 1, \dots, p$, respectively, where $0 \le \varphi_1^{\mathrm{SH}} \le \cdots \le \varphi_p^{\mathrm{SH}}$ and $0 \le \psi_p^{\mathrm{SH}} \le \cdots \le \psi_1^{\mathrm{SH}}$. For a detailed description of Stein's isotonic regression, see, for example, Lin and Perlman (1985). Now we define the Stein–Haff type estimator for ξ as

$$\hat{\xi}^{\mathrm{SH}} = B^{-1}\Phi^{\mathrm{SH}}BX_1 + B^{-1}(I - \Phi^{\mathrm{SH}})BX_2,$$

with $\Phi^{\mathrm{SH}} = \mathrm{diag}(\phi_1^{\mathrm{SH}}, \dots, \phi_p^{\mathrm{SH}})$ and $\phi_i^{\mathrm{SH}} = \psi_i^{\mathrm{SH}}/(\varphi_i^{\mathrm{SH}} + \psi_i^{\mathrm{SH}})$, whenever $i = 1, \dots, p$. It is easy to see that in this case we have $\phi_1^{\mathrm{SH}} \ge \cdots \ge \phi_p^{\mathrm{SH}} \ge 0$.

Remark. Let p be fixed and let n tend to ∞. Then for a fixed set of parameters Σ_1, Σ_2 we have

$$\alpha_i^{\mathrm{SH}} \sim n, \qquad \beta_i^{\mathrm{SH}} \sim n.$$

Hence $\phi_i^{\mathrm{SH}} \sim f_i$ for $1 \le i \le p$. This implies that $\hat{\xi}^{\mathrm{SH}} \sim \hat{\xi}^{\mathrm{LS}}$. This is a reassuring result since $\hat{\xi}^{\mathrm{LS}}$ is asymptotically efficient under these conditions.

Remark. By adapting the ideas of Haff (1982) to this problem, it is shown in Section 5 that the Stein–Haff type estimator also has an approximate formal Bayes interpretation.

3. Monte Carlo study. Due to the rather complicated nature of its construction, we have not been able to give an analytical treatment of the risk performance of the estimator $\hat{\xi}^{\mathrm{SH}}$. We shall instead observe the risk behavior of this estimator via a Monte Carlo study. For the simulations, the following variance-reduction technique is used: First let

$$\hat{\xi} = B^{-1}\Phi BX_1 + B^{-1}(I - \Phi)BX_2,$$

where $\Phi(F) = \mathrm{diag}(\phi_1, \dots, \phi_p)$, $B(S_1 + S_2)B' = I$, $BS_2B' = F = \mathrm{diag}(f_1, \dots, f_p)$ with $f_1 \ge \cdots \ge f_p$.

Proposition 2. *With the above notation,*

$$R(\hat{\xi}; \xi, \Sigma_1, \Sigma_2) = p + E(\hat{\xi} - \hat{\xi}^{\mathrm{BE}})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \hat{\xi}^{\mathrm{BE}}).$$

Proof. We observe that

$$\begin{aligned} R(\hat{\xi}; \xi, \Sigma_1, \Sigma_2) &= E\big[(\hat{\xi} - \hat{\xi}^{\mathrm{BE}})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \hat{\xi}^{\mathrm{BE}}) \\ &\qquad + (\hat{\xi}^{\mathrm{BE}} - \xi)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi}^{\mathrm{BE}} - \xi) \\ &\qquad + 2(\hat{\xi} - \hat{\xi}^{\mathrm{BE}})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi}^{\mathrm{BE}} - \xi)\big] \\ &= p + E(\hat{\xi} - \hat{\xi}^{\mathrm{BE}})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \hat{\xi}^{\mathrm{BE}}). \end{aligned}$$

Since $E(\hat{\xi} - \hat{\xi}^{\mathrm{BE}})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi}^{\mathrm{BE}} - \xi) = 0$, the last equality follows from Proposition 1. □

LEMMA 1. *With the above notation, let*

$$Y = (\Sigma_1^{-1} + \Sigma_2^{-1})^{1/2}(\hat{\xi}^{\mathrm{BE}} - \xi),$$
$$Z = (\Sigma_1^{-1} + \Sigma_2^{-1})^{1/2}(\hat{\xi} - \hat{\xi}^{\mathrm{BE}}).$$

Then Y *and* Z *are conditionally independent given* S_1, S_2.

PROOF. It is straightforward to show that $E^{S_1, S_2}YZ' = 0$. Also we note that conditional on S_1 and S_2, both Y and Z follow a multivariate normal distribution. Hence we conclude that Y and Z are conditionally independent. □

PROPOSITION 3. *With the above notation, let*

$$V = (\hat{\xi} - \hat{\xi}^{\mathrm{BE}})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \hat{\xi}^{\mathrm{BE}}),$$
$$W = (\hat{\xi} - \xi)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \xi).$$

Then $\mathrm{Var}[E(V|S_1, S_2)] \le \mathrm{Var}(V) \le \mathrm{Var}(W)$.

PROOF. From Lemma 1 and the proof of Proposition 2, we have

$$\begin{aligned}\mathrm{Var}(W) &= E\,\mathrm{Var}(W|S_1, S_2) + \mathrm{Var}\,E(W|S_1, S_2)\\ &\ge E\,\mathrm{Var}(V|S_1, S_2) + \mathrm{Var}\,E(V|S_1, S_2)\\ &= \mathrm{Var}(V)\\ &\ge \mathrm{Var}\,E(V|S_1, S_2).\end{aligned}$$

This completes the proof. □

For the simulations, independent standard normal variates are generated by the IMSL subroutine DRNNOA and the eigenvalue decomposition uses the IMSL subroutine DEVCSF. Also we take $p = 5$, $n = 7, 15, 30$ and $p = 10$, $n = 12, 25, 50$. Since Propositions 2 and 3 show that $\mathrm{Var}\,E(V|S_1, S_2) \le \mathrm{Var}(W)$ and $E(W) = p + E(V)$, a simulation is done to estimate $E(V)$; this will permit the estimation of $R(\hat{\xi}; \xi, \Sigma_1, \Sigma_2)$. We do this by computing the mean $\overline{V}$ of $E(V|S_1, S_2)$ based on 500 independent replications. Then we calculate the average loss of the estimator $\hat{\xi}$ by the formula $\overline{L} = p + \overline{V}$. Tables 1 and 2 give the average losses and their standard deviations of the estimators $\hat{\xi}^{\mathrm{LS}}$ and $\hat{\xi}^{\mathrm{SH}}$. We further observe from Proposition 1 that the risk of $\hat{\xi}^{\mathrm{BE}}$ is equal to p. This serves as a lower bound on the risks of these estimators.

Compared with the naive Monte Carlo, this variance-reduction technique on the average reduces the estimated standard deviations here by a factor of roughly around 10. Also we note 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 positive correlation between the average losses of these estimators. Since we are more interested in the relative risk ordering of these estimators, we conclude that

TABLE 1
Average losses of estimators for the common mean (estimated standard errors are in parentheses)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\xi}^{LS}$	$\hat{\xi}^{SH}$
$p = 5$ $n = 7$		
$(1, 1, 1, 1, 1)$	8.12 (0.06)	6.34 (0.04)
$(10, 0.1, 0.1, 0.1, 0.1)$	8.95 (0.14)	7.77 (0.11)
$(10, 10, 10, 0.1, 0.1)$	8.58 (0.09)	8.29 (0.08)
$(10, 1, 1, 1, 0.1)$	8.21 (0.07)	7.49 (0.05)
$(10, 10, 1, 0.1, 0.1)$	8.44 (0.08)	8.21 (0.07)
$(20, 5, 1, 0.5, 0.05)$	8.36 (0.08)	8.03 (0.07)
$(10^{10}, 5, 1, 0.5, 10^{-10})$	8.42 (0.08)	8.17 (0.08)
$(5, 2, 1, 0.5, 0.2)$	8.23 (0.07)	7.23 (0.05)
$(16, 8, 4, 2, 1)$	9.09 (0.11)	7.10 (0.08)
$(10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10})$	7.66 (0.13)	7.66 (0.13)
$(10^{8}, 10^{4}, 1, 10^{-4}, 10^{-8})$	8.57 (0.11)	8.60 (0.11)
$p = 5$ $n = 15$		
$(1, 1, 1, 1, 1)$	6.12 (0.02)	5.36 (0.01)
$(10, 0.1, 0.1, 0.1, 0.1)$	6.04 (0.02)	5.78 (0.02)
$(10, 10, 10, 0.1, 0.1)$	6.13 (0.02)	6.03 (0.02)
$(10, 1, 1, 1, 0.1)$	6.14 (0.02)	5.96 (0.02)
$(10, 10, 1, 0.1, 0.1)$	6.13 (0.02)	6.11 (0.02)
$(20, 5, 1, 0.5, 0.05)$	6.14 (0.02)	6.12 (0.02)
$(10^{10}, 5, 1, 0.5, 10^{-10})$	6.13 (0.02)	6.09 (0.02)
$(5, 2, 1, 0.5, 0.2)$	6.14 (0.02)	5.94 (0.01)
$(16, 8, 4, 2, 1)$	6.07 (0.02)	5.64 (0.01)
$(10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10})$	5.71 (0.02)	5.71 (0.02)
$(10^{8}, 10^{4}, 1, 10^{-4}, 10^{-8})$	6.11 (0.02)	6.12 (0.02)
$p = 5$ $n = 30$		
$(1, 1, 1, 1, 1)$	5.53 (0.01)	5.16 (0.01)
$(10, 0.1, 0.1, 0.1, 0.1)$	5.42 (0.01)	5.33 (0.01)
$(10, 10, 10, 0.1, 0.1)$	5.49 (0.01)	5.44 (0.01)
$(10, 1, 1, 1, 0.1)$	5.52 (0.01)	5.43 (0.01)
$(10, 10, 1, 0.1, 0.1)$	5.50 (0.01)	5.49 (0.01)
$(20, 5, 1, 0.5, 0.05)$	5.51 (0.01)	5.51 (0.01)
$(10^{10}, 5, 1, 0.5, 10^{-10})$	5.50 (0.01)	5.50 (0.01)
$(5, 2, 1, 0.5, 0.2)$	5.52 (0.01)	5.49 (0.01)
$(16, 8, 4, 2, 1)$	5.43 (0.01)	5.34 (0.01)
$(10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10})$	5.30 (0.01)	5.30 (0.01)
$(10^{8}, 10^{4}, 1, 10^{-4}, 10^{-8})$	5.48 (0.01)	5.48 (0.01)

TABLE 2
Average losses of estimators for the common mean (estimated standard errors are in parentheses)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\xi}^{LS}$	$\hat{\xi}^{SH}$
$p = 10$ $n = 12$		
(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)	17.71 (0.08)	12.33 (0.06)
(10, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)	22.87 (0.24)	15.73 (0.16)
(10, 10, 10, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)	19.33 (0.16)	16.99 (0.13)
(10, 10, 10, 10, 10, 0.1, 0.1, 0.1, 0.1, 0.1)	18.18 (0.12)	17.08 (0.11)
(10, 10, 10, 10, 10, 10, 10, 10, 0.1, 0.1)	20.00 (0.18)	16.21 (0.12)
(10, 9/2, 8/3, 7/4, 6/5, 5/6, 4/7, 3/8, 2/9, 1/10)	17.90 (0.09)	15.02 (0.06)
(10, 10, 10, 1, 1, 1, 1, 0.1, 0.1, 0.1)	18.01 (0.10)	16.18 (0.07)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	23.09 (0.28)	14.73 (0.14)
(32, 16, 8, 4, 2, 1, 1/2, 1/4, 1/8, 1/16)	18.02 (0.10)	16.48 (0.08)
(720, 360, 120, 30, 6, 1, 1/6, 1/30, 1/120, 1/360)	18.23 (0.12)	18.17 (0.12)
($10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}$)	16.06 (0.62)	16.06 (0.62)
($10^{10}, 10^{10}, 10^{10}, 10^{10}, 10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}$)	18.29 (0.14)	18.29 (0.14)
($10^5, 10^4, 10^3, 10^2, 10, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$)	18.25 (0.13)	18.25 (0.13)
$p = 10$ $n = 25$		
(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)	12.67 (0.02)	10.48 (0.01)
(10, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)	12.35 (0.03)	11.18 (0.02)
(10, 10, 10, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)	12.58 (0.03)	12.05 (0.02)
(10, 10, 10, 10, 10, 0.1, 0.1, 0.1, 0.1, 0.1)	12.65 (0.03)	12.31 (0.02)
(10, 10, 10, 10, 10, 10, 10, 10, 0.1, 0.1)	12.50 (0.03)	11.69 (0.02)
(10, 9/2, 8/3, 7/4, 6/5, 5/6, 4/7, 3/8, 2/9, 1/10)	12.68 (0.02)	12.08 (0.02)
(10, 10, 10, 1, 1, 1, 1, 0.1, 0.1, 0.1)	12.69 (0.02)	12.30 (0.02)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	12.13 (0.04)	11.07 (0.02)
(32, 16, 8, 4, 2, 1, 1/2, 1/4, 1/8, 1/16)	12.67 (0.02)	12.47 (0.02)
(720, 360, 120, 30, 6, 1, 1/6, 1/30, 1/120, 1/360)	12.61 (0.03)	12.62 (0.03)
($10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}$)	10.98 (0.02)	10.98 (0.02)
($10^{10}, 10^{10}, 10^{10}, 10^{10}, 10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}$)	12.56 (0.03)	12.56 (0.03)
($10^5, 10^4, 10^3, 10^2, 10, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$)	12.59 (0.03)	12.59 (0.03)
$p = 10$ $n = 50$		
(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)	11.18 (0.01)	10.17 (0.01)
(10, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)	10.81 (0.01)	10.47 (0.01)
(10, 10, 10, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)	11.07 (0.01)	10.88 (0.01)
(10, 10, 10, 10, 10, 0.1, 0.1, 0.1, 0.1, 0.1)	11.16 (0.01)	11.02 (0.01)
(10, 10, 10, 10, 10, 10, 10, 10, 0.1, 0.1)	10.96 (0.01)	10.71 (0.01)
(10, 9/2, 8/3, 7/4, 6/5, 5/6, 4/7, 3/8, 2/9, 1/10)	11.17 (0.01)	11.05 (0.01)
(10, 10, 10, 1, 1, 1, 1, 0.1, 0.1, 0.1)	11.17 (0.01)	11.00 (0.01)
(512, 256, 128, 64, 32, 16, 8, 4, 2, 1)	10.69 (0.01)	10.49 (0.01)
(32, 16, 8, 4, 2, 1, 1/2, 1/4, 1/8, 1/16)	11.16 (0.01)	11.15 (0.01)
(720, 360, 120, 30, 6, 1, 1/6, 1/30, 1/120, 1/360)	11.14 (0.01)	11.14 (0.01)
($10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}$)	10.41 (0.01)	10.41 (0.01)
($10^{10}, 10^{10}, 10^{10}, 10^{10}, 10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}$)	11.13 (0.01)	11.13 (0.01)
($10^5, 10^4, 10^3, 10^2, 10, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$)	11.13 (0.01)	11.13 (0.01)

the estimated standard deviation (as given in Tables 1 and 2) is probably a conservative indicator of the variability of the relative magnitude of the average losses.

We shall now summarize the results of this numerical study:

1. The risk of the estimator $\hat{\xi}^{\text{SH}}$ compares very favorably with that of $\hat{\xi}^{\text{LS}}$. In particular, when p and n are of comparable magnitude, significant savings in risk are achieved in most parts of the parameter space. This is most evident when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are close together. For example, in the case of $p = 10$, $n = 12$ and the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ being all equal to 1, approximately 30% savings in risk is achieved with the use of $\hat{\xi}^{\text{SH}}$ over that of $\hat{\xi}^{\text{LS}}$.
2. However, when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are far apart, there does not appear to be any significant difference in risk among the two estimators $\hat{\xi}^{\text{LS}}$ and $\hat{\xi}^{\text{SH}}$.
3. There are a few cases where the average loss of $\hat{\xi}^{\text{SH}}$ exceeds that of $\hat{\xi}^{\text{LS}}$. These occur when the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ are extremely far apart; for example, when $p = 5$, $n = 7$, $\Sigma_2\Sigma_1^{-1} = \text{diag}(10^8, 10^4, 1, 10^{-4}, 10^{-8})$. Even though the differences between these average losses in these situations are well within the estimated standard deviations, due to the high positive correlation between the average losses of $\hat{\xi}^{\text{LS}}$ and $\hat{\xi}^{\text{SH}}$, it seems likely that $\hat{\xi}^{\text{SH}}$ does not dominate $\hat{\xi}^{\text{LS}}$. However, in these cases, the average loss of $\hat{\xi}^{\text{SH}}$ exceeds that of $\hat{\xi}^{\text{LS}}$ by at most 1% of the average loss of $\hat{\xi}^{\text{LS}}$. Since this is usually acceptable in most applied work, this study indicates that the Stein–Haff type estimator offers an attractive alternative to the usual generalized least squares estimator.
4. For a fixed set of parameters $(\xi, \Sigma_1, \Sigma_2)$, the study also shows that the savings in risk of $\hat{\xi}^{\text{SH}}$ over $\hat{\xi}^{\text{LS}}$ increase with p and decrease with n.

4. Unbiased estimate of risk. We shall state the normal and Wishart identities. These identities are crucial in the derivation of the unbiased estimate of risk of an equivariant estimator in the problem that we are considering.

First we need some additional definitions. 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.

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

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

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

The normal identity was first proved by Stein (1973). Let $\mathscr{S}_p$ denote the set of $p \times p$ positive definite matrices. Also we write for $1 \le i, j \le p$,

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

where δ_{ij} denotes the Kronecker delta.

THEOREM 3 (Wishart identity). *Let $X = (X_1, \dots, X_n)$ be a $p \times n$ random matrix, with the X_k independently normally distributed p-dimensional random vectors with mean 0 and unknown covariance matrix Σ. We suppose $n \ge p$. 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 $S = (s_{ij}) = XX'$, we have*

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

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. Now in what follows, let $X_1 \sim N_p(\xi, \Sigma_1)$, $S_1 \sim W_p(\Sigma_1, n)$, $X_2 \sim N_p(\xi, \Sigma_2)$ and $S_2 \sim W_p(\Sigma_2, n)$. For $i = 1, 2$ and $1 \le j, k \le p$, we write

$$\nabla^{(i)} = \left(\nabla_1^{(i)}, \dots, \nabla_p^{(i)}\right)', \qquad \tilde{\nabla}^{(i)} = \left(\tilde{\nabla}_{jk}^{(i)}\right)_{p\times p},$$

where

$$x_j^{(i)} = (X_i)_j, \qquad \nabla_j^{(i)} = \partial/\partial x_j^{(i)},$$

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

We observe that there exists $B \in GL(p, R)$ such that $BS_1B' = I - F$ and $BS_2B' = F$, where $F = \operatorname{diag}(f_1, \dots, f_p)$ and $f_1 \ge \cdots \ge f_p$. We shall now compute the partial derivatives of B, B^{-1} and F with respect to S_1 and S_2.

LEMMA 2. *Let $X_i \sim N_p(\xi, \Sigma_i)$ and $S_i \sim W_p(\Sigma_i, n)$, $i = 1, 2$. Then with F, $B = (b_{il})$ and $B^{-1} = (b^{il})$ as defined above, we have*

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

$$\tilde{\nabla}_{jk}^{(2)} f_i = (1-f_i) b_{ij} b_{ik},$$

$$\tilde{\nabla}_{jk}^{(1)} b_{il} = -\frac{1}{2} b_{il} b_{ij} b_{ik} - \frac{1}{2} \sum_{k' \ne i} b_{k'l}\left(b_{ij} b_{k'k} + b_{ik} b_{k'j}\right) \frac{f_i}{f_i - f_{k'}},$$

$$\tilde{\nabla}_{jk}^{(2)} b_{il} = -\frac{1}{2} b_{il} b_{ij} b_{ik} + \frac{1}{2} \sum_{k' \ne i} b_{k'l}\left(b_{ij} b_{k'k} + b_{ik} b_{k'j}\right) \frac{1-f_i}{f_i - f_{k'}},$$

$$\tilde{\nabla}_{jk}^{(1)} b^{il} = \frac{1}{2} b^{il} b_{lj} b_{lk} - \frac{1}{2} \sum_{i' \ne l} b^{ii'}\left(b_{i'j} b_{lk} + b_{l'k} b_{lj}\right) \frac{f_{i'}}{f_l - f_{i'}},$$

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

The proof of this lemma uses the calculus on eigenstructure techniques of Stein (1975, 1977a) and Haff (1982, 1988). We refer the reader to Loh (1988) for the proof. Now we shall compute the unbiased estimate of risk of a subclass of equivariant estimators for ξ of the form

$$\hat{\xi} = B^{-1}\Phi BX_1 + B^{-1}(I-\Phi)BX_2,$$

where $\Phi = \operatorname{diag}(\phi_1, \dots, \phi_p)$ depends only on F. It is clear that estimators of this kind are unbiased. First we need a couple of rather technical lemmas.

LEMMA 3. *With the notation of Theorem* 1, *for* $\Phi = \Phi(F)$ *we have*

$$\nabla^{(1)\prime}[B^{-1}(I-\Phi)B(X_2 - X_1)] = -p + \sum_i \phi_i,$$

$$\nabla^{(2)\prime}[B^{-1}\Phi B(X_1 - X_2)] = -\sum_i \phi_i.$$

PROOF. We observe that

$$\begin{aligned}
&\nabla^{(1)\prime}[B^{-1}(I-\Phi)B(X_2 - X_1)] \\
&= \sum_i \sum_j \frac{\partial}{\partial x_i^{(1)}}[B^{-1}(I-\Phi)B]_{ij}(X_2 - X_1)_j \\
&= -\sum_i [B^{-1}(I-\Phi)B]_{ii} \\
&= -p + \sum_i \phi_i.
\end{aligned}$$

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

LEMMA 4. *With the notation of Theorem* 1, *for* $\Phi = \Phi(F)$ *we have*

1. $$\begin{aligned}
&\operatorname{tr} \tilde{\nabla}^{(1)}\{[B^{-1}(I-\Phi)B(X_2 - X_1)][B^{-1}(I-\Phi)B(X_2 - X_1)]'\} \\
&= \sum_i \Bigg\{[B(X_1 - X_2)]_i^2(1-\phi_i)^2 \sum_{j \neq i} \frac{f_j}{f_j - f_i} \\
&\qquad + 2[B(X_1 - X_2)]_i^2(1-\phi_i) f_i \frac{\partial \phi_i}{\partial f_i} \\
&\qquad - \sum_{j \neq i} [B(X_1 - X_2)]_j^2 (1-\phi_i)(1-\phi_j)\frac{f_i}{f_i - f_j}\Bigg\},
\end{aligned}$$

2. $$\begin{aligned}
&\operatorname{tr} \tilde{\nabla}^{(2)}\{[B^{-1}\Phi B(X_1 - X_2)][B^{-1}\Phi B(X_1 - X_2)]'\} \\
&= \sum_i \Bigg\{[B(X_1 - X_2)]_i^2 \phi_i^2 \sum_{j \neq i} \frac{1-f_j}{f_i - f_j} \\
&\qquad + 2[B(X_1 - X_2)]_i^2 \phi_i (1-f_i)\frac{\partial \phi_i}{\partial f_i} \\
&\qquad - \sum_{j \neq i} [B(X_1 - X_2)]_j^2 \phi_i \phi_j \frac{1-f_i}{f_j - f_i}\Bigg\}.
\end{aligned}$$

PROOF. First we observe that

$$\begin{aligned}
&\{\tilde{\nabla}^{(1)}[B^{-1}(I-\Phi)B(X_2-X_1)]\}_i \\
&\quad = \sum_{j,k,l} \tilde{\nabla}_{ij}^{(1)}[b^{jl}(1-\phi_l)b_{lk}](X_2-X_1)_k \\
&\quad = \sum_{k,l}\left[\frac{1}{2}b_{li}b_{lk}(1-\phi_l)\sum_{i'\neq l}\frac{f_{i'}}{f_{i'}-f_l} + b^{jl}b_{lk}\sum_m\left(b_{mi}b_{mj}f_m\frac{\partial\phi_l}{\partial f_m}\right)\right. \\
&\qquad\qquad \left. -\frac{1}{2}\sum_{k'\neq l} b_{k'k}b_{k'i}(1-\phi_l)\frac{f_l}{f_l-f_{k'}}\right](X_2-X_1)_k \qquad (3)\\
&\quad = \sum_l \left\{\frac{1}{2}b_{li}[B(X_2-X_1)]_l(1-\phi_l)\sum_{i'\neq l}\frac{f_{i'}}{f_{i'}-f_l} + b_{li}[B(X_2-X_1)]_l f_l\frac{\partial\phi_l}{\partial f_l}\right. \\
&\qquad\qquad \left. -\frac{1}{2}\sum_{k'\neq l} b_{k'i}[B(X_2-X_1)]_{k'}(1-\phi_l)\frac{f_l}{f_l-f_{k'}}\right\}.
\end{aligned}$$

Here the second last equality follows from Lemma 2. Hence we have

$$\begin{aligned}
&\operatorname{tr}\tilde{\nabla}^{(1)}\{[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]'\} \\
&\quad = 2\operatorname{tr}[\tilde{\nabla}^{(1)}B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]' \\
&\quad = 2\sum_i [\tilde{\nabla}^{(1)}B^{-1}(I-\Phi)B(X_2-X_1)]_i[B^{-1}(I-\Phi)B(X_2-X_1)]_i \\
&\quad = \sum_i \left\{[B(X_1-X_2)]_i^2(1-\phi_i)^2\sum_{j\neq i}\frac{f_j}{f_j-f_i}\right. \\
&\qquad\qquad +2[B(X_1-X_2)]_i^2(1-\phi_i)f_i\frac{\partial\phi_i}{\partial f_i} \\
&\qquad\qquad \left. -\sum_{j\neq i}[B(X_1-X_2)]_j^2(1-\phi_i)(1-\phi_j)\frac{f_i}{f_i-f_j}\right\}.
\end{aligned}$$

The last equality follows from (3). The second part of this lemma can be proved similarly. □

Now we shall prove the main result of this section.

THEOREM 4. *Let $\hat{\xi}$ be an estimator for ξ, where*

$$\hat{\xi}(X_1, X_2, S_1, S_2) = B^{-1}\Phi BX_1 + B^{-1}(I-\Phi)BX_2,$$

$\Phi = \Phi(F) = \operatorname{diag}(\phi_1, \dots, \phi_p)$, $B(S_1+S_2)B' = I$, $BS_2B' = F = \operatorname{diag}(f_1, \dots, f_p)$ *with* $f_1 \geq \cdots \geq f_p$. *Suppose Φ satisfies the conditions of the normal and*

Wishart identities, then the risk of $\hat{\xi}$ *is given by*

$$R(\hat{\xi};\xi,\Sigma_1,\Sigma_2) = E\Bigg\{\sum_i [B(X_1 - X_2)]_i^2 \Bigg[\frac{n-p-1}{f_i}\phi_i^2 + 4(1-f_i)\phi_i\frac{\partial\phi_i}{\partial f_i} + 2\sum_{j\neq i}\phi_i(\phi_i-\phi_j)\frac{1-f_j}{f_i-f_j} + \frac{n-p-1}{1-f_i}(1-\phi_i)^2 + 4f_i(1-\phi_i)\frac{\partial\phi_i}{\partial f_i} + 2\sum_{j\neq i}(1-\phi_i)(\phi_i-\phi_j)\frac{f_j}{f_i-f_j}\Bigg]\Bigg\}.$$

PROOF. We observe that

$$\begin{aligned} R(\hat{\xi};\xi,\Sigma_1,\Sigma_2) &= E(\hat{\xi}-\xi)'(\Sigma_1^{-1}+\Sigma_2^{-1})(\hat{\xi}-\xi) \\ &= E\operatorname{tr}\{2p + 2(X_1-\xi)'\Sigma_1^{-1}B^{-1}(I-\Phi)B(X_2-X_1) \\ &\quad + \Sigma_1^{-1}[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]' \\ &\quad + 2(X_2-\xi)'\Sigma_2^{-1}B^{-1}\Phi B(X_1-X_2) \\ &\quad + \Sigma_2^{-1}[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]'\}. \end{aligned}$$

Since Φ satisfies the conditions of the normal and Wishart identities, we have

$$\begin{aligned} R(\hat{\xi};\xi,\Sigma_1,\Sigma_2) &= E\operatorname{tr}\{2p + 2\nabla^{(1)\prime}[B^{-1}(I-\Phi)B(X_2-X_1)] \\ &\quad + (n-p-1)S_1^{-1}[B^{-1}(I-\Phi)B(X_2-X_1)] \\ &\quad \times [B^{-1}(I-\Phi)B(X_2-X_1)]' \\ &\quad + 2\tilde{\nabla}^{(1)}\{[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]'\} \\ &\quad + 2\nabla^{(2)\prime}[B^{-1}\Phi B(X_1-X_2)] \\ &\quad + (n-p-1)S_2^{-1}[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]' \\ &\quad + 2\tilde{\nabla}^{(2)}\{[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]'\}\}. \end{aligned}$$

Now it follows from Lemmas 3 and 4 that

$$\begin{aligned} R(\hat{\xi};\xi,\Sigma_1,\Sigma_2) &= E\Bigg\{\sum_i [B(X_1 - X_2)]_i^2 \Bigg[\frac{n-p-1}{f_i}\phi_i^2 + 4(1-f_i)\phi_i\frac{\partial\phi_i}{\partial f_i} \\ &\quad + 2\sum_{j\neq i}\phi_i(\phi_i-\phi_j)\frac{1-f_j}{f_i-f_j} + \frac{n-p-1}{1-f_i}(1-\phi_i)^2 \\ &\quad + 4f_i(1-\phi_i)\frac{\partial\phi_i}{\partial f_i} + 2\sum_{j\neq i}(1-\phi_i)(\phi_i-\phi_j)\frac{f_j}{f_i-f_j}\Bigg]\Bigg\}. \end{aligned}$$

This completes the proof. □

5. Derivation of the Stein–Haff type estimator. In view of Section 2.2, it suffices to derive (2). Let $\hat{\xi}$ be an equivariant estimator for ξ, where

$$\hat{\xi}(X_1, X_2, S_1, S_2) = B^{-1}\Phi B X_1 + B^{-1}(I - \Phi) B X_2,$$

$\Phi = \Phi(F) = \operatorname{diag}(\phi_1, \dots, \phi_p)$, $B(S_1 + S_2)B' = I$, $BS_2B' = F = \operatorname{diag}(f_1, \dots, f_p)$ with $f_1 \ge \cdots \ge f_p$.

LEMMA 5. *With the above notation, for* $1 \le i \le p$ *we have*

$$E^{S_1, S_2}[B(X_1 - X_2)]_i^2 = E^{S_1, S_2}[B(\Sigma_1 + \Sigma_2)B']_{ii},$$

where E^{S_1, S_2} *denotes conditional expectation given* S_1, S_2.

PROOF. We observe that

$$\begin{aligned} E^{S_1, S_2}[B(X_1 - X_2)]_i^2 &= E^{S_1, S_2}\sum_{j,k} b_{ij}b_{ik}\big(x_j^{(1)} - x_j^{(2)}\big)\big(x_k^{(1)} - x_k^{(2)}\big) \\ &= E^{S_1, S_2}\sum_{j,k} b_{ij}b_{ik}(\Sigma_1 + \Sigma_2)_{jk} \\ &= E^{S_1, S_2}[B(\Sigma_1 + \Sigma_2)B']_{ii}. \qquad \square \end{aligned}$$

For $i = 1, 2$, by replacing Σ_i in the above lemma with its maximum likelihood estimator S_i/n, we get the following approximation:

$$E^{S_1, S_2}[B(X_1 - X_2)]_i^2 \approx E^{S_1, S_2}[B(S_1 + S_2)B']_{ii}/n = 1/n. \tag{4}$$

Next it follows from Theorem 4 that the unbiased estimate of the risk of $\hat{\xi}$ can be expressed as

$$\begin{aligned} \hat{R} = \sum_i [B(X_1 - X_2)]_i^2 \Bigg[&\frac{n-p-1}{f_i}\phi_i^2 + 4(1 - f_i)\phi_i\frac{\partial \phi_i}{\partial f_i} \\ &+ 2\sum_{j \ne i} \phi_i(\phi_i - \phi_j)\frac{1 - f_j}{f_i - f_j} + \frac{n-p-1}{1 - f_i}(1 - \phi_i)^2 \\ &+ 4f_i(1 - \phi_i)\frac{\partial \phi_i}{\partial f_i} + 2\sum_{j \ne i}(1 - \phi_i)(\phi_i - \phi_j)\frac{f_j}{f_i - f_j}\Bigg]. \end{aligned}$$

From (4), we observe that $\hat{R}$ can be approximated by

$$\begin{aligned} \hat{R} \approx \frac{1}{n}\sum_i \Bigg\{ &\frac{n-p-1}{f_i}\phi_i^2 + 4(1 - f_i)\phi_i\left[f_i\frac{\partial}{\partial f_i}\left(\frac{\phi_i}{f_i}\right) + \frac{\phi_i}{f_i}\right] \\ &+ 2\sum_{j \ne i}\phi_i(\phi_i - \phi_j)\frac{1 - f_j}{f_i - f_j} + \frac{n-p-1}{1 - f_i}(1 - \phi_i)^2 \\ &+ 4f_i(1 - \phi_i)\left[(1 - f_i)\frac{\partial}{\partial(1 - f_i)}\left(\frac{1 - \phi_i}{1 - f_i}\right) + \frac{1 - \phi_i}{1 - f_i}\right] \\ &+ 2\sum_{j \ne i}(1 - \phi_i)(\phi_i - \phi_j)\frac{f_j}{f_i - f_j}\Bigg\} \\ = \tilde{R}, \quad &\text{say}. \end{aligned} \tag{5}$$

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

$$\check{R} = \frac{1}{n}\sum_i \Bigg[\frac{n-p-1}{f_i}\phi_i^2 + \frac{1-f_i}{f_i}4\phi_i^2 + 2\sum_{j\neq i}\phi_i(\phi_i-\phi_j)\frac{1-f_j}{f_i-f_j} + \frac{n-p-1}{1-f_i}(1-\phi_i)^2 + \frac{f_i}{1-f_i}4(1-\phi_i)^2 + 2\sum_{j\neq i}(1-\phi_i)(\phi_i-\phi_j)\frac{f_j}{f_i-f_j}\Bigg].$$

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

$$\begin{aligned} 0 &= \frac{\partial \check{R}}{\partial \phi_i} \\ &= \frac{n-p-1}{f_i}\phi_i + \frac{1-f_i}{f_i}4\phi_i + 2\phi_i\sum_{j\neq i}\frac{1-f_j}{f_i-f_j} - \sum_{j\neq i}\phi_j \\ &\quad - \frac{n-p-1}{1-f_i}(1-\phi_i) - \frac{f_i}{1-f_i}4(1-\phi_i) \\ &\quad + 2(1-\phi_i)\sum_{j\neq i}\frac{f_j}{f_i-f_j} + \sum_{j\neq i}(1-\phi_j). \end{aligned} \tag{6}$$

Solving (6) would lead us to an estimator for ξ. However, we observe that in the problem of estimating a normal covariance matrix, Haff (1988) achieves a slight improvement over the Stein (1975) estimator by using an approximate formal Bayes method. Motivated by this, we shall introduce a prior on the parameter space and use (6) as a guide to the choice of m which is defined below.

Put a prior distribution on the parameter space $\{(\xi, \Sigma_1, \Sigma_2)\colon \xi \in R^p$ and Σ_1, Σ_2 being positive definite matrices$\}$ and let $m(F)$ denote the marginal density of F. It follows from (5) that the unbiased estimate of the risk of $\hat{\xi}$ can be approximated by $\tilde{R}$. This implies that the approximate average risk of this estimator is

$$\int \hat{G}(f_1,\dots,f_p;\phi_1,\dots,\phi_p;\partial\phi_1/\partial f_1,\dots,\partial\phi_p/\partial f_p)\,dF,$$

where $\hat{G} = m\tilde{R}$. The solution of the Euler–Lagrange equations minimizes the above integral. These equations are given by

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

where $\hat{G}_{\phi_i} = \partial \hat{G}/\partial \phi_i$, etc. Evaluating the above equations gives for $i = 1, \ldots, p$,

$$0 = \frac{1}{f_i}\left[(n-p-1)\phi_i + 2\phi_i \sum_{j\neq i} \frac{f_i(1-f_j)}{f_i - f_j} + 2f_i\phi_i - f_i \sum_{j\neq i} \phi_j - 2f_i(1-f_i)\phi_i \frac{\partial \log m}{\partial f_i}\right]$$

$$(7) \qquad - \frac{1}{1-f_i}\left[(n-p-1)(1-\phi_i) - 2(1-\phi_i)\sum_{j\neq i}\frac{(1-f_i)f_j}{f_i-f_j} + 2(1-f_i)(1-\phi_i) - (1-f_i)\sum_{j\neq i}(1-\phi_j) + 2f_i(1-f_i)(1-\phi_i)\frac{\partial \log m}{\partial f_i}\right].$$

Next we set

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

This choice of m is motivated by the following reasons: (i) by equivariance, m should be a symmetric function of the f_i's and $(1 - f_i)$'s, (ii) by that choice of m, (7) should resemble (6) as much as possible. This leaves us with

$$0 = \frac{1}{f_i}\left[(n-p-1)\phi_i + 2\phi_i \sum_{j\neq i} \frac{f_i(1-f_j)}{f_i - f_j} + 2(1-f_i)\phi_i - f_i \sum_{j\neq i} \phi_j\right]$$
$$- \frac{1}{1-f_i}\left[(n-p-1)(1-\phi_i) - 2(1-\phi_i)\sum_{j\neq i}\frac{(1-f_i)f_j}{f_i-f_j} + 2f_i(1-\phi_i) - (1-f_i)\sum_{j\neq i}(1-\phi_j)\right].$$

For computational simplicity, we ignore the last term in each of the square brackets. We observe that these terms do not contribute significantly to the r.h.s. of the above equation. Solving for ϕ_i, we get

$$\phi_i = [\beta_i^{\mathrm{SH}}/(1-f_i)]/\{[\alpha_i^{\mathrm{SH}}/f_i] + [\beta_i^{\mathrm{SH}}/(1-f_i)]\},$$

where

$$\alpha_i^{\mathrm{SH}} = n - p - 1 + 2(1 - f_i) + 2\sum_{j \neq i} \frac{f_i(1 - f_j)}{f_i - f_j},$$

$$\beta_i^{\mathrm{SH}} = n - p - 1 + 2f_i - 2\sum_{j \neq i} \frac{(1 - f_i) f_j}{f_i - f_j}.$$

6. Final remarks. The main obstacle in using the unbiased estimate of risk to get good estimators is the fact that risk is a "smooth version" of the unbiased estimate of risk and hence the unbiased estimate of risk does not reflect exactly the behavior of the risk. This is self-evident since we need to integrate the unbiased estimate of risk to get to the risk. Thus except for special cases, proving theoretical dominance over the usual estimator is generally not possible with this method, assuming of course that the usual estimator is inadmissible. However, as this paper indicates, the unbiased estimate of risk does possess a good deal of useful information. If this is exploited carefully, possibly with the help of a computer, one can obtain an alternative estimator which competes very favorably against the usual estimator.

Acknowledgments. This paper is a revised version of a part of my Ph.D. dissertation at Stanford University. I would like to express my deep gratitude to my advisor, Professor Charles Stein, whose constant support, encouragement and guidance made this dissertation possible. I would like to thank Professors Art Cohen and L. R. Haff for alerting me to relevant literature which I had overlooked and for sending me copies of their work. Finally, I am extremely grateful to two anonymous referees for their detailed comments and suggestions.

REFERENCES

BROWN, L. D. and COHEN, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information. *Ann. Statist.* **2** 963–976.

CHIOU, W. and COHEN, A. (1985). On estimating a common multivariate normal mean vector. *Ann. Inst. Statist. Math.* **37** 499–506.

COHEN, A. and SACKROWITZ, H. B. (1974). On estimating the common mean of two normal distributions. *Ann. Statist.* **2** 1274–1282.

GRAYBILL, F. A. and DEAL, R. B. (1959). Combining unbiased estimators. *Biometrics* **15** 543–550.

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

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

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

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

KHATRI, C. G. and SHAH, K. R. (1974). Estimation of location parameters from two linear models under normality. *Comm. Statist. B—Simulation Comput.* **3** 647–665.

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

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

SHINOZAKI, N. (1978). A note on estimating the common mean of k normal distributions and the Stein problem. *Comm. Statist. A—Theory Methods* **7** 1421–1432.

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

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

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

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

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

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907