Sankhyā : The Indian Journal of Statistics
1985, Volume **47**, Series A, Pt. 2, pp. 247-254.

EMPIRICAL BAYES ESTIMATORS OF NORMAL COVARIANCE MATRIX

By DIVAKAR SHARMA and K. KRISHNAMOORTHY

Indian Institute of Technology, Kanpur, India

SUMMARY. Let $\boldsymbol{S}$ follow a Wishart distribution, with n degrees of freedom and parameter $\boldsymbol{\Sigma}$. Haff (1979b, 1980) has proposed estimators of $\boldsymbol{\Sigma}$ which dominate the best multiples of $\boldsymbol{S}$ under certain loss functions. We consider some other loss functions, show that these estimators are empirical Bayes and compare one of them with the best multiples of $\boldsymbol{S}$.

1. Introduction

Let $\boldsymbol{S}$ be a $p \times p$ random matrix following a Wishart distribution with its probability density proportional to

$$|\boldsymbol{S}|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}}. \quad \dots \quad (1.1)$$

We are interested in the estimation of $\boldsymbol{\Sigma}$ under the loss functions

$$L_1(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}}) = \operatorname{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}-\boldsymbol{I})^2,$$

$$L_2(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}}) = \frac{\operatorname{tr}(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})^2\boldsymbol{S}^{-1}}{\operatorname{tr}\boldsymbol{\Sigma}},$$

and

$$L_3(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}}) = \frac{\operatorname{tr}(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})^2\boldsymbol{\Sigma}^{-1}}{\operatorname{tr}\boldsymbol{\Sigma}}.$$

The loss function L_1 was first used by Selliah (1964) and is essentially "squared error", L_2 is analogous to one considered by Efron and Morris (1976) for estimating $\boldsymbol{\Sigma}^{-1}$ and L_3 is similar to the loss function in Haff (1979a, 1981). Whereas following Selliah (1964), loss function L_1 has received the attention of many authors, for example Haff (1980), Sharma (1980),, Sugiura and Fujimoto (1982), L_2 and L_3 have probably not appeared in the literature earlier. It is well known that $\boldsymbol{S}/(n+p+1)$ is the best multiple of $\boldsymbol{S}$ under L_1. Various better estimators have been obtained; Selliah (1964) showed that the best lower triangular equivariant estimator is such an estimator while Haff (1980) proved this property for an estimator of the form

$$a\boldsymbol{S}+b(\operatorname{tr}\boldsymbol{S}^{-1})^{-1}\boldsymbol{I}. \quad \dots \quad (1.2)$$

AMS (1980) *subject classification :* 62F10.

Key words and phrases : Empirical Bayes estimators, Normal covariance matrix.

In fact, Haff proved that the estimator (1.2) is better than the best multiple aS also under Stein's loss function

$$L_4(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}) = \operatorname{tr} \hat{\mathbf{\Sigma}}\, \mathbf{\Sigma}^{-1} - \log \det. \hat{\mathbf{\Sigma}}\, \mathbf{\Sigma}^{-1} - p$$

if $b \leqslant 2(p-1)/n^2$. Haff gave an empirical Bayes interpretation to the estimator. We point out in Section 2 that the estimator under the loss functions L_2 and L_3 also could be viewed as empirical Bayes and compare it with the best multiples of S in Section 3. In Section 2, we also show that Haff's (1979b) estimator $a\mathbf{S}+b|\mathbf{S}|^{1/p}\mathbf{I}$, which dominates the best multiple aS under the loss function $\hat{\Sigma}_{i \leqslant j} (\hat{\sigma}_{ij}-\sigma_{ij})^2 q_{ij}$, where $\hat{\mathbf{\Sigma}} = (\hat{\sigma}_{ij})$ and $q_{ij} > 0$, is empirical Bayes under the loss functions L_1, L_2 and L_3.

2. Haff's estimators as empirical Bayes estimators

We need the following result due to Eaton (1970) to exhibit the empircal Bayes nature of Haff's (1979b, 1980) estimators.

Let $(\mathscr{X}, \mathscr{B}, \mathscr{P})$ be a probability space with $P \in \mathscr{P} = \{P_\theta : \mathbf{\theta} \in \Theta\}$, a family of probability distributions. Consider a mapping $g : \Theta \rightarrow (\mathscr{A}, (\cdot, \cdot))$, where $(\mathscr{A}, (\cdot, \cdot))$ is a finite dimensional inner product space with $(\cdot, \cdot)$ the inner product. Let π be a prior distribution on Θ and the problem be one of estimating $g(\theta)$ with a loss function

$$L(\mathbf{\theta}, a) = (g(\mathbf{\theta})-a, \tau(\mathbf{\theta})(g(\mathbf{\theta})-a))$$

where $a \in \mathscr{A}$ is the action and $\tau(\theta)$ is a positive definite linear transformation, for each $\theta \in \Theta$, from $\mathscr{A}$ into $\mathscr{A}$. Then the Bayes estimator and the Bayes risk are given by

Lemma 2.1 (Eaton, 1970) : *The Bayes estimator* $\hat{a}(X)$ *of* $g(\mathbf{\theta})$ *is*

$$[E_{\theta \mid \boldsymbol{x}}\, \tau(\mathbf{\theta})]^{-1} E_{\theta \mid \boldsymbol{x}}\, (\tau(\mathbf{\theta}) g(\mathbf{\theta}))$$

with its Bayes risk equal to

$$E_\theta(g(\mathbf{\theta}), \tau(\mathbf{\theta})g(\mathbf{\theta})) - E_{\boldsymbol{X}}(E_{\mathbf{\theta} \mid \boldsymbol{X}}\, \tau(\mathbf{\theta})g(\mathbf{\theta}), \hat{a}(X)),$$

where $E_{\boldsymbol{X}}$, $E_{\mathbf{\theta}}$, *and* $E_{\mathbf{\theta} \mid \boldsymbol{X}}$ *denote the expectations with respect to the marginal distribution of* X, *the prior distribution of* $\mathbf{\theta}$ *and the posterior distribution of* $\mathbf{\theta}$ *respectively.*

Designate the distribution assumption (1.1) by $\mathbf{S} \frown W_p(\mathbf{\Sigma}, n)$. Assume that $\mathbf{\Sigma}^{-1} = \mathbf{A}$ (say) $\frown W_p(\mathbf{I}/c, k)$. Then $\mathbf{A} \mid S \frown W_p((\mathbf{S}+c\mathbf{I})^{-1}, m)$, where

$m = n+k$. Define an inner product on the vector space of all the $p \times p$ matrices as $(\boldsymbol{C}, \boldsymbol{D}) = \operatorname{tr} \boldsymbol{C}\boldsymbol{D}'$ then, L_1 can be written as

$$L_1(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}}) = (\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}, \tau(\boldsymbol{\Sigma})(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})),$$

where $\tau(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}$. Here $\otimes$ stands for the Kronecker product, that is, for $p \times p$ matrices $\boldsymbol{C}$, $\boldsymbol{D}$ and $\boldsymbol{H}$,

$$(\boldsymbol{C} \otimes \boldsymbol{D})\boldsymbol{H} = \boldsymbol{C}\boldsymbol{H}\boldsymbol{D}'. \qquad \ldots \quad (2.2)$$

The definition (2.2) is equivalent to

$$\boldsymbol{C} \otimes \boldsymbol{D} = \begin{pmatrix} d_{11}\boldsymbol{C} & \ldots & d_{1p}\boldsymbol{C} \\ d_{p1}\boldsymbol{C} & \ldots & d_{pp}\boldsymbol{C} \end{pmatrix} \qquad \ldots \quad (2.3)$$

with the understanding that in carrying out the operation $(\boldsymbol{C} \otimes \boldsymbol{D})\boldsymbol{H}$ when $\boldsymbol{C} \otimes \boldsymbol{D}$ is defined by (2.3), the $p \times p$ matrix $\boldsymbol{H}$ is rearranged as a p^2-dimensional column vector with the original $(i+1)$-th column of $\boldsymbol{H}$ put below its i-th column.

Clearly, $\tau(\boldsymbol{\Sigma})$ is linear and is positive definite with respect to the inner product (2.1). Thus, for the loss L_1, from Lemma 2.1, the Bayes estimator is

$$[E_{\boldsymbol{A}|\boldsymbol{S}} A \otimes A]^{-1} E_{\boldsymbol{A}|\boldsymbol{S}} (A \otimes A) A^{-1} = [E_{\boldsymbol{A}|\boldsymbol{S}} A \otimes A]^{-1} E_{\boldsymbol{A}|\boldsymbol{S}} A. \ \ldots \quad (2.4)$$

To evaluate (2.4), let $\boldsymbol{M} = (\boldsymbol{S}+c\boldsymbol{I})^{-1}$, $\boldsymbol{M}^{1/2}$ a symmetric matrix satisfying $(\boldsymbol{M}^{1/2})^2 = \boldsymbol{M}$; and $\boldsymbol{B} = \boldsymbol{M}^{-1/2}\boldsymbol{A}\boldsymbol{M}^{-1/2}$. Then $\boldsymbol{B}|\boldsymbol{S} \frown W_p(\boldsymbol{I}, m)$ and

$$\begin{aligned} E_{\boldsymbol{A}|\boldsymbol{S}} (\boldsymbol{A} \otimes \boldsymbol{A}) &= E_{\boldsymbol{B}|\boldsymbol{S}}[(\boldsymbol{M}^{1/2}\boldsymbol{B}\boldsymbol{M}^{1/2}) \otimes (\boldsymbol{M}^{1/2}\boldsymbol{B}\boldsymbol{M}^{1/2})] \\ &= E_{\boldsymbol{B}|\boldsymbol{S}}[(\boldsymbol{M}^{1/2} \otimes \boldsymbol{M}^{1/2})(\boldsymbol{B} \otimes \boldsymbol{B})(\boldsymbol{M}^{1/2} \otimes \boldsymbol{M}^{1/2})] \\ &= (\boldsymbol{M}^{1/2} \otimes \boldsymbol{M}^{1/2}) E_{\boldsymbol{B}|\boldsymbol{S}}(\boldsymbol{B} \otimes \boldsymbol{B})(\boldsymbol{M}^{1/2} \otimes \boldsymbol{M}^{1/2}). \end{aligned}$$

Now, for any orthogonal $\boldsymbol{R}$;

$$\begin{aligned} E_{\boldsymbol{B}|\boldsymbol{S}}(\boldsymbol{B} \otimes \boldsymbol{B}) &= E_{\boldsymbol{B}|\boldsymbol{S}}[(\boldsymbol{R}\boldsymbol{B}\boldsymbol{R}') \otimes (\boldsymbol{R}\boldsymbol{B}\boldsymbol{R}')] \\ &= (\boldsymbol{R} \otimes \boldsymbol{R})[E_{\boldsymbol{B}|\boldsymbol{S}}(\boldsymbol{B} \otimes \boldsymbol{B})](\boldsymbol{R}' \otimes \boldsymbol{R}')\boldsymbol{I}. \end{aligned}$$

Hence,

$$\begin{aligned} [E_{\boldsymbol{B}|\boldsymbol{S}}(\boldsymbol{B} \otimes \boldsymbol{B})]^{-1}\boldsymbol{I} &= \boldsymbol{N} \text{ (say)} \\ &= (\boldsymbol{R} \otimes \boldsymbol{R})[E_{\boldsymbol{B}|\boldsymbol{S}}(\boldsymbol{B} \otimes \boldsymbol{B})]^{-1}(\boldsymbol{R}' \otimes \boldsymbol{R}')\boldsymbol{I} \\ &= \boldsymbol{R}\boldsymbol{N}\boldsymbol{R}'. \end{aligned}$$

Choosing $\boldsymbol{R}$ to be a diagonal matrix with the diagonal elements 1 or -1, it can be easily seen that $\boldsymbol{N} = \alpha_m \boldsymbol{I}$ for some real α_m. Thus,

$$[E_{\boldsymbol{B}|\boldsymbol{S}}(\boldsymbol{B} \otimes \boldsymbol{B})]^{-1}\boldsymbol{I} = \alpha_m \boldsymbol{I}. \qquad \ldots \quad (2.5)$$

To calculate α_m, we use the (2.3) definition of $\otimes$, let $\boldsymbol{I}$ in (2.5) be a p^2-dimensional vector and rewrite (2.5) as

$$\alpha_m[E_{\boldsymbol{B}|\boldsymbol{S}}(\boldsymbol{B}\otimes\boldsymbol{B})] = \boldsymbol{I}. \qquad \ldots \quad (2.6)$$

Equating the first components of the p^2-dimensional vectors on the two sides of the relation (2.6), we get

$$\alpha_m(m^2+2m)+\alpha_m(p-1)m = 1$$

so that $\alpha_m = (m^2+m(p+1))^{-1}$. Hence the Bayes estimator is

$$(\boldsymbol{S}+c\boldsymbol{I})/(m+p+1).$$

Similarly, L_3 can be written as

$$L_3(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}}) = \left(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}, \frac{1}{\operatorname{tr}\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\right)$$

so that $\tau(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1}/\operatorname{tr}\boldsymbol{\Sigma}$ is positive definite and linear. Now, for large n, $\tau(\boldsymbol{\Sigma})$ is close to $n\boldsymbol{\Sigma}^{-1}/\operatorname{tr} S$ and so using Lemma 2.1, the Bayes estimator is approximated by

$$\left[E_{\boldsymbol{\Sigma}|\boldsymbol{S}}\frac{\boldsymbol{\Sigma}^{-1}}{\operatorname{tr} S}\right]^{-1} E_{\boldsymbol{\Sigma}|\boldsymbol{S}}\left(\frac{\boldsymbol{\Sigma}^{-1}}{\operatorname{tr} S}\boldsymbol{\Sigma}\right) = [E_{\boldsymbol{\Sigma}|\boldsymbol{S}}\boldsymbol{\Sigma}^{-1}]^{-1} = \frac{\boldsymbol{S}+c\boldsymbol{I}}{m}. \quad \ldots \quad (2.7)$$

As $L_2(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}})/n$, for large n, is close to $L_3(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}})$, an approximate Bayes estimator for L_2 also is $(S+cI)/m$.

Whether we consider the Bayes estimator $(\boldsymbol{S}+c\boldsymbol{I})/(m+p+1)$ under L_1, or the approximate Bayes estimator $(\boldsymbol{S}+c\boldsymbol{I})/m$ under L_2 and L_3, the main idea is to estimate c and arrive at an estimator of $\boldsymbol{\Sigma}$ which is better than the best multiple of S in the frequency sense.

From the fact that the marginal density of $\boldsymbol{S}$ is proportional to

$$|\boldsymbol{S}|^{n-p-1)/2}\, c^{pk/2}/|\boldsymbol{S}+c\boldsymbol{I}|^{m/2}. \qquad \ldots \quad (2.8)$$

Haff (1980) has proved that $g(\boldsymbol{S})\propto(\operatorname{tr}\boldsymbol{S}^{-1})^{-1}$ is a generalized maximum likelihood estimator of c and so the estimator (1.2) with an appropriate a can be described as empirical Bayes under L_i $(i = 1, 2, 3)$.

From (2.8), we also notice that $E|S|^{\beta/2}\propto c^{\beta p/2}$. Take $\beta = 2/p$, then $E|\boldsymbol{S}|^{1/p}\propto c$. Substituting the estimator of c in the Bayes or approximate Bayes estimator of $\boldsymbol{\Sigma}$ obtained above, Haff's (1979b) estimator $a\boldsymbol{S}+b|\boldsymbol{S}|^{1/p}\boldsymbol{I}$ also can be seen to be empirical Bayes under L_i $(i = 1, 2, 3)$.

The results in the next section concern the estimator (1.2). We shall use the notation R_i for the risk under the loss function L_i.

3. THE ESTIMATORS $a\mathbf{S}+b(\text{tr } \mathbf{S}^{-1})^{-1}\mathbf{I}$

We first state Haff's (1979b) identity which we need for proving the dominance results in this section.

For any $p\times p$ matrix $\mathbf{M} = (m_{ij}(\mathbf{S}))$ of the $p\times p$ matrix $\mathbf{S}$, define $\mathbf{M}_{(c)} = (m'_{ij})$ where

$$m'_{ij} = \begin{cases} m_{ij} & \text{for } i = j \\ cm_{ij} & \text{for } i \neq j, \end{cases}$$

and $\mathbf{D}^*\mathbf{M} = \Sigma\Sigma\partial m_{ij}/\partial s_{ij}$. Suppose now $\mathbf{S} \frown W_p(\mathbf{\Sigma}, n)$, $h(\mathbf{S})$ is real-valued and $\mathbf{T} = T(\mathbf{S})$ is a $p\times p$ matrix then

$$E[h(\mathbf{S})\text{tr } \mathbf{T}\mathbf{\Sigma}^{-1}] = 2Eh(\mathbf{S})\mathbf{D}^*\mathbf{T}_{(1/2)}+2E \text{ tr } \left[\frac{\partial h(S)}{\partial S} \mathbf{T}_{(1/2)} \right]$$

$$+(n-p-1)E[h(\mathbf{S})\text{tr } \mathbf{S}^{-1}\mathbf{T}]. \qquad \ldots \quad (3.1)$$

For the validity of (3.1), the reader is referred to Haff (1979b).

Next, we show that the best choice of a among estimators aS is n^{-1} under L_2.

Theorem 3.1 : *For the loss function L_2, the best multiple of $\mathbf{S}$ is $\mathbf{S}/n$. It is also the best unbiased estimator of $\mathbf{\Sigma}$.*

Proof : The risk of $a\mathbf{S}$ is

$$R_2(\mathbf{\Sigma}, a\mathbf{S}) = E \text{ tr}(a\mathbf{S}-\mathbf{\Sigma})^2S^{-1}/\text{tr } \mathbf{\Sigma} = a^2n-2a+(n-p-1)^{-1}$$

and it has minimum value $(p+1)n^{-1}(n-p-1)^{-1}$ at $a = n^{-1}$. The completeness of $\mathbf{S}$ and the convexity of the loss function (see, for example, the proof for L_1 in Sharma (1980)) imply that $\mathbf{S}/n$ is the best unbiased estimator of $\mathbf{\Sigma}$.

The following theorem shows that it is possible to choose b such that $a\mathbf{S}+b(\text{tr } \mathbf{S}^{-1})^{-1}\mathbf{I}$ is better than aS under L_2.

Theorem 3.2 : *Let $\hat{\mathbf{\Sigma}}_2$ be given by* (1.2) *with*

$$a < (n-p+1)^{-1}, \; 0 < b \leqslant 2p[(n-p+1)^{-1}-a]. \qquad \ldots \quad (3.2)$$

Then $R_2(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_2) < R_2(\mathbf{\Sigma}, a\mathbf{S})$ for all $\mathbf{\Sigma}$.

Proof : Let $R_2(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_2)-R_2(\mathbf{\Sigma}, a\mathbf{S}) = \alpha_2(\mathbf{\Sigma})$, where

$$\alpha_2(\mathbf{\Sigma}) = \frac{b}{\text{tr } \mathbf{\Sigma}} E\left[\frac{b \text{ tr } \mathbf{S}^{-1}}{(\text{tr } \mathbf{S}^{-1})^2} + \frac{2ap}{\text{tr } \mathbf{S}^{-1}} - 2 \frac{\text{tr } \mathbf{S}^{-1}\mathbf{\Sigma}}{\text{tr } \mathbf{S}^{-1}} \right].$$

Now $\alpha_2(\mathbf{\Sigma}) < 0$ if

$$(b+2ap)E(\text{tr } \mathbf{S}^{-1})^{-1} - 2E(\text{tr } \mathbf{S}^{-1}\mathbf{\Sigma}/\text{tr } \mathbf{S}^{-1}) < 0. \qquad \ldots \quad (3.3)$$

However, (3.3) is true as proved below.

Using Haff's identity (3.1),

$$E(\text{tr } \mathbf{S}^{-1})^{-1} = \frac{2}{p} E(\text{tr } \mathbf{S}^{-1})^{-2}\text{tr}(\mathbf{S}^{-2}{}_{(2)}\mathbf{\Sigma}_{(1/2)} + \frac{n-p-1}{p} E \frac{\text{tr } \mathbf{S}^{-1}\mathbf{\Sigma}}{\text{tr } \mathbf{S}^{-1}}$$

$$= \frac{2}{p} E \frac{\text{tr } \mathbf{S}^{-2}\mathbf{\Sigma}}{(\text{tr } \mathbf{S}^{-1})^2} + \frac{n-p-1}{p} E \frac{\text{tr } \mathbf{S}^{-1}\mathbf{\Sigma}}{\text{tr } \mathbf{S}^{-1}}. \qquad \ldots \quad (3.4)$$

Hence, $\text{tr } \mathbf{S}^{-2}\mathbf{\Sigma} \leqslant \text{tr } \mathbf{S}^{-1}\mathbf{\Sigma} \text{ tr } \mathbf{S}^{-1}$ implies that

$$E(\text{tr } \mathbf{S}^{-1})^{-1} < \frac{n-p+1}{p} E \frac{\text{tr } \mathbf{S}^{-1}\mathbf{\Sigma},}{\text{tr } \mathbf{S}^{-1}}$$

so that $\alpha_2(\mathbf{\Sigma}) < 0$ if

$$(b+2ap)(n-p+1)/p - 2 \leqslant 0, \quad b > 0,$$

which is true from (3.2).

Under Stein's loss function L_4 also $\mathbf{S}/n$ is the best multiple of $\mathbf{S}$. Haff (Theorem 4.3, 1980) has given estimators better than $\mathbf{S}/n$ under L_4. His result can be combined with Theorem 3.1 to obtain estimators of Σ which are better than $\mathbf{S}/n$ both under L_2 and Stein's loss function.

Corollary 3.1 : *Let* $\hat{\mathbf{\Sigma}}_{2,4}$ *be the estimator* $\mathbf{S}/n + b(\text{tr } \mathbf{S}^{-1})^{-1}\mathbf{I}$ *with* $0 < b \leqslant 2(p-1)/n^2$. *Then, for all* $\mathbf{\Sigma}$, $R_2(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_{2,4}) < R_2(\mathbf{\Sigma}, \mathbf{S}/n)$ *and* $R_4(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_{2,4}) < R_4(\mathbf{\Sigma}, \mathbf{S}/n)$.

Theorems similar to 3.1 and 3.2, for the loss function L_3 are given below.

Theorem 3.3 : *For the loss function* L_3, *the best multiple of* S *is* $S/(n+p+1)$.

Proof : $R_3(\mathbf{\Sigma}, a\mathbf{S}) = a^2 E \text{ tr } \mathbf{S}^2\mathbf{\Sigma}^{-1} + \text{tr } \mathbf{\Sigma} - 2aE \text{ tr } \mathbf{S})/\text{tr } \mathbf{\Sigma}$. Since, from the identity (3.1) or otherwise, $E\mathbf{S}^2 = n(n+1)\mathbf{\Sigma}^2 + n\mathbf{\Sigma} \text{ tr } \mathbf{\Sigma}$, we have $R_3(\mathbf{\Sigma}, a\mathbf{S}) = a^2 n(n+p+1) - 2an + 1$, which is minimized at $a = (n+p+1)^{-1}$.

Theorem 3.4 : *Let* $\hat{\mathbf{\Sigma}}_3$ *be the estimator* (1.2) *with*

$$a < p/(np+2), \;\; 0 < b \leqslant 2[p - a(np+2)]/(n-p+3). \qquad \ldots \quad (3.5)$$

Then $R_3(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_3) < R_3(\mathbf{\Sigma}, a\mathbf{S})$ *for all* $\mathbf{\Sigma}$.

Proof : Let $R_3(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_3) - R_3(\mathbf{\Sigma}, a\mathbf{S}) = \alpha_3(\mathbf{\Sigma})$,

where $$\alpha_3(\mathbf{\Sigma}) = \frac{b}{\text{tr } \mathbf{\Sigma}} E\left[\frac{b \text{ tr } \mathbf{\Sigma}^{-1}}{(\text{tr } \mathbf{S}^{-1})^2} + \frac{2a \text{ tr } \mathbf{S}\mathbf{\Sigma}^{-1}}{\text{tr } \mathbf{S}^{-1}} - \frac{2p}{\text{tr } \mathbf{S}^{-1}}\right]$$

Now, from (3.1),

$$E\frac{\operatorname{tr}\mathbf{\Sigma}^{-1}}{(\operatorname{tr}\mathbf{S}^{-1})^2} = \frac{4E\operatorname{tr}(\mathbf{S}^{-2})_{(2)}}{(\operatorname{tr}\mathbf{S}^{-1})^3}+(n-p-1)E\frac{\operatorname{tr}\mathbf{S}^{-1}}{(\operatorname{tr}\mathbf{S}^{-1})^2} \leqslant (n-p+3)E(\operatorname{tr}\mathbf{S}^{-1})^{-1} \quad \ldots (3.6)$$

and

$$E\frac{\operatorname{tr}\mathbf{S\Sigma}^{-1}}{\operatorname{tr}\mathbf{S}^{-1}} = p(p+1)E(\operatorname{tr}\mathbf{S}^{-1})^{-1}+2E\frac{\operatorname{tr}(\mathbf{S}^{-2})_{(2)}\mathbf{S}_{(1/2)}}{(\operatorname{tr}\mathbf{S}^{-1})^2}+(n-p-1)pE(\operatorname{tr}\mathbf{S}^{-1})^{-1}$$

$$= (np+2)E(\operatorname{tr}\mathbf{S}^{-1})^{-1}. \quad \ldots (3.7)$$

From (3.6) and (3.7), it can be seen that $\alpha_3(\Sigma) < 0$ if

$$b(n-p+3)+2a(np+2)-2p \leqslant 0, \; b > 0,$$

which is true from (3.5).

Theorem 3.2 and 3.4 can be combined to get the following corollary.

Corollary 3.2 : *Let* $\hat{\mathbf{\Sigma}}_{2,3}$ *be the estimator* (1.2) *with*

$$a \leqslant (n+1)^{-1}, \; 0 < b \leqslant \min\left\{\frac{2[p-a(np+2)]}{n-p+3}, \frac{2p[1-a(n-p+1)]}{n-p+1}\right\},$$

then, for all $\mathbf{\Sigma}$,

$$R_2(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_{2,3}) < R_2(\Sigma, aS) \quad and \quad R_3(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_{2,3}) < R_3(\mathbf{\Sigma}, a\mathbf{S}).$$

Recall that $\mathbf{S}/(n+p+1)$ is the best multiple of $\mathbf{S}$ under Selliah's loss function L_1 also. Combining Theorem 3.4 and Haff's (1980) Theorem 4.5, we get estimators better than $\mathbf{S}/(n+p+1)$ both under L_3 and Selliah's loss function. This is stated as Corollary 3.3 below.

Corollary 3.3 : *Let* $\hat{\mathbf{\Sigma}}_{1,3}$ *be the estimator* $\mathbf{S}/(n+p+1)+b(tr\,\mathbf{S}^{-1})^{-1}I$, *with* $0 < b \leqslant 2(p-1)(n-p+3)^{-1}(n+p+1)^{-1}$, *then for all* $\mathbf{\Sigma}$,

$$R_1(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_{1,3}) < R_1(\mathbf{\Sigma}, \mathbf{S}/(n+p+1) \text{ and } R_3(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}_{1,3}) < R_3(\mathbf{\Sigma}, \mathbf{S}/(n+p+1)).$$

References

EATON, M. L. (1970): Some problems in covariance estimation. *Technical Report* No. 49, Department of Statistics, Stanford University.

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

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.

——— (1979b): An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, **9**, 531-544.

———— (1980): Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.*, **8**, 586-597.

———— (1981): Erratum, *J. Multivariate Anal.*, **11**, 608-609.

SELLIAH, J. B. (1964): Estimation and testing problems in Wishart distribution. *Technical Report No.* 10, Department of Statistics, Stanford University.

SHARMA, D. (1980): An estimator of normal covariance matrix, *Calcutta Statist. Assoc. Bull.*, **29**, 161-167.

SUGIURA, W. and FUJIMOTO, M. (1982): Asymptotic risk comparison of improved estimators for normal covariance matrix, *Tsukuba J. Math.*, **6**, 103-126.

Paper received : September, 1983.

Revised : July, 1984.