Journal of Mathematical Extension
Vol. 5, No. 2 (1), (2011), 31-46

Robust Empirical Bayes Estimation of the Elliptically Countoured Covariance Matrix

Z. Khodadadi
Islamic Azad University, Marvdasht-Branch

B. Tarami
Yasouj University, College of Sciences

Abstract. Let $\boldsymbol{S}$ be the matrix of residual sum of square in linear model $\boldsymbol{Y} = \boldsymbol{A\beta} + \mathbf{e}$, where the matrix of errors is distributed as elliptically contoured with unknown scale matrix $\boldsymbol{\Sigma}$. For Stein loss function, $L_1(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p$, and squared loss function, $L_2(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \mathbf{I})^2$, we offer empirical Bayes estimators of $\boldsymbol{\Sigma}$, which dominate any scalar multiple of $\boldsymbol{S}$, i.e., $a\boldsymbol{S}$, by an effective amount. In fact, this study somehow shows that improvement of the empirical Bayes estimators obtained under the normality assumption remains robust under elliptically contoured model.

AMS Subject Classification: 62C12.
Keywords and Phrases: Covariance matrix, elliptically contoured, empirical Bayes estimators.

1. Introduction

The problem of empirical Bayes (EB) estimation with the normal covariance matrix $\boldsymbol{\Sigma}$, was considered by [3], who proved that these estimators dominate all scalar multiples of the unbiased estimator. Our objective is to establish the dominance results by Haff [4] remains robust under the elliptically contoured distribution which we refer to it as ECD in this

Received November 2010; Final Revised February 2011

paper. Here, we consider the problem of estimation with the elliptically contoured covariance matrix $\mathbf{\Sigma}$; then we get the empirical Bayes estimators which dominate the usual unbiased estimators for each of two invariant loss functions L_1 and L_2. The dominance results under L_1 and L_2 were first offered by James and Stein. There are many studies to estimate a covariance matrix for the normality assumption; under loss functions L_1 and L_2, see [2,3], [6], [9] and [11], [8], [10]. The identity for the ECD which was derived by [7], known in the literature as the "Stein-Haff identity", is applied to compute risk functions.

Let $\mathbf{Y}$ be an $N \times p$ random matrix with multivariate linear model

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\beta} + \mathbf{e}$$

where $\mathbf{e}$ is an $N \times p$ matrix of random errors, $\mathbf{A}$ is a known full rank $N \times m$ matrix and $\boldsymbol{\beta}$ is an $m \times p$ matrix of unknown parameters. We assume that the error matrix $\mathbf{e}$ has an elliptical density

$$|\mathbf{\Sigma}|^{-\frac{N}{2}} f\big(\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{e}^t\mathbf{e})\big)$$

where $\mathbf{\Sigma}$ is a $p \times p$ unknown positive-definite matrix, $f(\cdot)$ is a differentiable and nonnegative real-value function. Here $|\mathbf{B}|$, $\mathrm{tr}(\mathbf{B})$ and $\mathbf{B}^t$ stand for the determinant, the trace and the transpose of a square matrix $\mathbf{B}$, respectively.

Let $\boldsymbol{S}$ be the matrix of residual sum of squares, i.e.,

$$\boldsymbol{S} = \boldsymbol{Y}^t(\boldsymbol{I}_N - \boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t)\boldsymbol{Y},$$

and let $n = N - m$. Under the elliptically assumption, the expected value for various functions of $\boldsymbol{S}$ have been derived; by [7].

Let $\hat{\mathbf{\Sigma}}$ be an estimator of $\mathbf{\Sigma}$. We assume that the loss function is

$$L_1(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) = \mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) - \log|\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}| - p,$$

or

$$L_2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) = \mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1} - \mathbf{I})^2,$$

and define the risk function by

$$R_i(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) = E[L_i(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma})|\mathbf{\Sigma}], \qquad i = 1, 2.$$

Let $\hat{\mathbf{\Sigma}}$ and $\hat{\mathbf{\Sigma}}_*$ be the competing estimators of $\mathbf{\Sigma}$, where $\hat{\mathbf{\Sigma}}$ dominates $\hat{\mathbf{\Sigma}}_*$ (mod L_i) if $R_i(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) \leqslant R_i(\hat{\mathbf{\Sigma}}_*, \mathbf{\Sigma})(\forall \mathbf{\Sigma})$.

According to the same notation used in [7], let $f^{(0)}(x) = f(x)$,

$$f^{(k+1)}(x) = \frac{1}{2}\int_x^{\infty} f^{(k)}(t)dt, \quad k = 0, 1$$

and

$$E_{\mathbf{\Sigma}}^{(k)}[v(\boldsymbol{S})] = \int v(\boldsymbol{S})\ |\mathbf{\Sigma}|^{-\frac{N}{2}} f^{(k)}\Big(\mathrm{tr}\big(\mathbf{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{A\beta})^t(\boldsymbol{y} - \boldsymbol{A\beta})\big)\Big)d\boldsymbol{y}$$

where $v(\boldsymbol{S})$ is an integrable function of $\boldsymbol{S}$. Also we use the transformation to polar coordinates to get,

$$E_{\mathbf{\Sigma}}^{(k)}[1] = \gamma(k) = \frac{2\pi^{NP/2}}{\Gamma(NP/2)}\int_0^{\infty} r^{Np-1} f^{(k)}(r^2)dr, \quad k = 0, 1, 2 \qquad (1)$$

and assume that $\gamma(i) < \infty$, for more details see [7].

Following [3], the empirical Bayes estimators have the form

$$\hat{\mathbf{\Sigma}} = a[\boldsymbol{S} + ut(u)\boldsymbol{C}] \qquad (2)$$

[5] where $t(\cdot)$ is a nonnegative and non-increasing function, $\mathbf{C}$ is an arbitrary positive definite matrix, $u = (\mathrm{tr}\boldsymbol{S}^{-1})^{-1}$ and

$$0 \leqslant a \leqslant max\{\frac{1}{n\gamma(1)}, \frac{\gamma(1)}{(n+p+1)\gamma(2)}\}.$$

Without loss of generality, we assume $\boldsymbol{C} = \boldsymbol{I}$. It should be noted that for $t \equiv 0$, we have the obvious estimators, the scalar multiples of $\boldsymbol{S}$.

1.1 Synopsis

The consideration of the scalar multiples of $\boldsymbol{S}$ shows that the best estimator (mod L_1) is the unbiased estimator

$$\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{n\gamma(1)}\,\boldsymbol{S}$$

and the best estimator (mod L_2) is

$$\hat{\boldsymbol{\Sigma}}_2 = \frac{\gamma(1)}{(n+p+1)\gamma(2)}\,\boldsymbol{S}.$$

The main result of this paper deals with the EB estimators(2). Furthermore, for each loss function, there are conditions under which they dominate the best scalar multiple of $\boldsymbol{S}$.

1.2 Stein-Haff Identity and Its Application

Let $T(\boldsymbol{S}) = (t_{ij}(\boldsymbol{S}))$ be a $p \times p$ matrix whose elements are functions of $\boldsymbol{S} = (s_{ij})$. Denote

$$\{D_{\boldsymbol{S}}T(\boldsymbol{S})\}_{ij} = \sum_{a=1}^{p} \frac{1}{2}(1+\delta_{ia})\frac{\partial t_{aj}(\boldsymbol{S})}{\partial s_{ia}} \tag{3}$$

where δ_{ia} is Kronecker's delta. From Lemma 1 in [11], for suitable choice of a matrix $T(\boldsymbol{S})$, the Stein-Haff identity is given by

$$E^{(k)}_{\boldsymbol{\Sigma}}\Big[\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}T(\boldsymbol{S})\}\Big] = E^{(k+1)}_{\boldsymbol{\Sigma}}\Big[(n-p-1)\ \mathrm{tr}\{\boldsymbol{S}^{-1}T(\boldsymbol{S})\} + 2\ \mathrm{tr}D_{\boldsymbol{S}}T(\boldsymbol{S})\Big] \tag{4}$$

and from (3) and (4) for a real valued function $h(\boldsymbol{S})$, we observe that

$$E^{(k)}_{\boldsymbol{\Sigma}}\Big[\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}h(\boldsymbol{S})\big)\Big] = E^{(k+1)}_{\boldsymbol{\Sigma}}\Big[(n-p-1)\ h(\boldsymbol{S})\ \mathrm{tr}(\boldsymbol{S}^{-1}) + \mathrm{tr}\big(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\big)\Big] \tag{5}$$

First, we apply (4) and (5) to calculate risk function R_1. It appears that $\alpha_1(\boldsymbol{\Sigma}) = R_1(\hat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}) - R_1(\hat{\boldsymbol{\Sigma}}_1,\boldsymbol{\Sigma})$ has terms under the unusual expectation of the form $h(\boldsymbol{S})\mathrm{tr}(\boldsymbol{\Sigma}^{-1})$. Theorem 2.3 gives conditions under which $\alpha_1(\boldsymbol{\Sigma}) \leqslant 0$ $(\forall \boldsymbol{\Sigma})$. Since $\alpha_2(\boldsymbol{\Sigma}) = R_2(\hat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}) - R_2(\hat{\boldsymbol{\Sigma}}_2,\boldsymbol{\Sigma})$ has terms which are quadratic in $\boldsymbol{\Sigma}^{-1}$, calculating R_2 is more difficult than R_1.

2. Main Results

The following four theorems are the main results of this paper. The proofs are postponed to Section 3.

Theorem 2.1. *Under the loss function L_1, the best estimator of the form $\hat{\mathbf{\Sigma}}_a = a\boldsymbol{S}$ is given by $a = \frac{1}{n\gamma(1)}$.*

Theorem 2.2. *Under the loss function L_2, the best estimator of the form $\hat{\mathbf{\Sigma}}_a = a\boldsymbol{S}$ is given by $a = \frac{\gamma(1)}{(n+p+1)\gamma(2)}$.*

Our main result concern the EB estimators (2). For comparing $\hat{\mathbf{\Sigma}}$ with $\hat{\mathbf{\Sigma}}_i, i = 1, 2$, a is replaced by $\frac{1}{n\gamma(1)}$ and $\frac{\gamma(1)}{(n+p+1)\gamma(2)}$.

Theorem 2.3. *Let $\hat{\mathbf{\Sigma}}$ is given by (2), with*

i) $a = \frac{1}{n\gamma(1)}$,

ii) $u = (\mathrm{tr}\boldsymbol{S}^{-1})^{-1}$

iii) *t is a constant,* $0 \leqslant t \leqslant \frac{2(n\gamma(0)+p-n)}{n\gamma(0)}$.

Then $\hat{\mathbf{\Sigma}}$ dominates $\hat{\mathbf{\Sigma}}_1$ (mod L_1), i.e., $R_1(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) \leqslant R_1(\hat{\mathbf{\Sigma}}_1, \mathbf{\Sigma})$ $(\forall \mathbf{\Sigma})$. An optimal value of t is $\frac{n\gamma(0)+p-n}{n\gamma(0)}$ (as seen from the proof).

Theorem 2.4. *Let $\hat{\mathbf{\Sigma}}$ is given by (2), with*

i) $a = \frac{\gamma(1)}{(n+p+1)\gamma(2)}$,

ii) $u = (\mathrm{tr}\ \boldsymbol{S}^{-1})^{-1}$

iii) *t is a constant,* $0 \leqslant t \leqslant \frac{2\left[(n-2p+1)-p(n-1)(n-p)\right]}{p(n-p-1)(n-p+1)}$.

Then $\hat{\mathbf{\Sigma}}$ dominates $\hat{\mathbf{\Sigma}}_2$ (mod L_2), i.e., $R_2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) \leqslant R_2(\hat{\mathbf{\Sigma}}_2, \mathbf{\Sigma})$ $(\forall \mathbf{\Sigma})$.

The choice $\frac{(n-2p+2)-p(n-1)(n-p)}{p(n-p-1)(n-p+1)}$ of t is the optimal choice.

The latter calculations depends on the following lemmas.

Lemma 2.5. [Hisayuki Tsukuma 2005] *Let* $\mathbf{Q}$ *be a* $p \times p$ *matrix of constants. Under the conditions of Lemma 2.1. in [5](see appendix), we have*

i) $E_{\boldsymbol{\Sigma}}^{(0)}[\boldsymbol{S}] = n\gamma(1)\boldsymbol{\Sigma}$,

ii) $E_{\boldsymbol{\Sigma}}^{(0)}[\boldsymbol{S}\mathbf{Q}\boldsymbol{S}] = \gamma(2)\big\{n^2\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Sigma} + n\boldsymbol{\Sigma}\mathbf{Q}^t\boldsymbol{\Sigma} + n\,\mathrm{tr}(\mathbf{Q}\boldsymbol{\Sigma})\boldsymbol{\Sigma}\big\}$.

Lemma 2.6. *Let* $\mathbf{F}$ *be a* $p \times p$ *matrix-valued function of* $\boldsymbol{S} = (s_{ij})$ *and* $\phi(\boldsymbol{S})$ *be a scalar function of* $\boldsymbol{S}$. *Then, we have*

i) $D_{\boldsymbol{S}}(\phi\,\mathbf{F}) = \frac{1}{2}(\frac{\partial\phi}{\partial\boldsymbol{S}} \cdot \mathbf{F}) + \phi\, D_{\boldsymbol{S}}\mathbf{F}$,

and for a matrix $\mathbf{Q}_{p\times p}$ *of constants,*

ii) $D_{\boldsymbol{S}}(\boldsymbol{S}\mathbf{Q}) = (\mathbf{Q} + p\,\mathbf{Q}^t)/2$,

iii) $D_{\boldsymbol{S}}(\boldsymbol{S}^{-2}\,\mathbf{Q}) = -\boldsymbol{S}^{-3}\mathbf{Q} - \big\{\mathrm{tr}(\boldsymbol{S}^{-1})\boldsymbol{S}^{-2}\mathbf{Q} + \mathrm{tr}(\boldsymbol{S}^{-2})\boldsymbol{S}^{-1}\mathbf{Q}\big\}/2$,

iv) $D_{\boldsymbol{S}}(\boldsymbol{S}^{-1}\mathbf{Q}) = -\big\{\boldsymbol{S}^{-2}\mathbf{Q} + \mathrm{tr}(\boldsymbol{S}^{-1})\boldsymbol{S}^{-1}\mathbf{Q}\big\}/2$.

Proof.(i) For a matrix $F(\boldsymbol{S})_{p\times p}$ and a scalar $\phi(\boldsymbol{S})$, from (3) we have

$$\begin{aligned}
[D_{\boldsymbol{S}}\phi\mathbf{F}]_{ij} &= \sum_a \frac{1}{2}(1+\delta_{ia})\frac{\partial}{\partial s_{ia}}\,(\phi\mathbf{F})_{aj} \\
&= \sum_a \frac{1}{2}(1+\delta_{ia})\left\{\frac{\partial\phi}{\partial s_{ia}}(\mathbf{F})_{aj} + \frac{\partial(\mathbf{F})_{aj}}{\partial s_{ia}}\cdot\phi\right\} \\
&= \sum_a \frac{1}{2}(1+\delta_{ia})(\frac{\partial\phi}{\partial\boldsymbol{S}})_{ia}\,(\mathbf{F})_{aj} \\
&+ \phi\sum_a \frac{1}{2}(1+\delta_{ia})\frac{\partial(\mathbf{F})_{aj}}{\partial s_{ia}} \\
&= \frac{1}{2}\big(\frac{\partial\phi}{\partial\boldsymbol{S}}\cdot\mathbf{F}\big)_{ij} + \phi\,(D_{\boldsymbol{S}}\mathbf{F})_{ij}
\end{aligned}$$

which gives Lemma 3.2 (i). □

The following properties of the operator $D_{\boldsymbol{S}}$ (see the definition in (3)) are required for computations.

Proof.(ii) For a matrix Q of constant, to derive Lemma 3.2 (ii), we note that

$$\begin{aligned}[D_{\boldsymbol{S}}\boldsymbol{S}\mathbf{Q}]_{ij} &= \frac{1}{2}\sum_a (1+\delta_{ia})\frac{\partial}{\partial s_{ia}}(\boldsymbol{S}\mathbf{Q})_{aj} \\ &= \frac{1}{2}\sum_a \left\{\delta_{ia}\mathbf{Q}_{aj} + \delta_{aa}\mathbf{Q}^t_{ij}\right\} \\ &= \frac{1}{2}(\mathbf{IQ})_{ij} + \frac{P}{2}\ \mathbf{Q}^t_{ij}.\end{aligned}$$

Proof.(iii)Applying

$$\frac{\partial(\boldsymbol{S}^{-1})_{kl}}{\partial s_{ij}} = \{-(\boldsymbol{S}^{-1})_{ki}(\boldsymbol{S}^{-1})_{jl} - (\boldsymbol{S}^{-1})_{li}(\boldsymbol{S}^{-1})_{jk}\},$$

we obtain

$$\begin{aligned}[D_{\boldsymbol{S}}\boldsymbol{S}^{-2}\mathbf{Q}]_{ij} &= \sum_{a,k}\frac{1}{2}(1+\delta_{ia})\frac{\partial}{\partial s_{ia}}\Big[(\boldsymbol{S}^{-1})_{ak}\ (\boldsymbol{S}^{-1}\mathbf{Q})_{kj}\Big] \\ &= -\frac{1}{2}\sum_{a,k}(\boldsymbol{S}^{-1})_{ia}\ (\boldsymbol{S}^{-1})_{ak}\ (\boldsymbol{S}^{-1}\mathbf{Q})_{kj} \\ &- \frac{1}{2}\Big(\sum_k (\boldsymbol{S}^{-1})_{ik}\ (\boldsymbol{S}^{-1}\mathbf{Q})_{kj}\Big)\ \Big(\sum_a (\boldsymbol{S}^{-1})_{aa}\Big) \\ &- \frac{1}{2}\Big(\sum_{a,k}(\boldsymbol{S}^{-1})_{ak}\ (\boldsymbol{S}^{-1})_{ki}\Big)\ \Big(\sum_m (\boldsymbol{S}^{-1})_{am}\ \mathbf{Q}_{mj}\Big) \\ &- \frac{1}{2}\Big(\sum_{a,k}(\boldsymbol{S}^{-1})_{ak}\ (\boldsymbol{S}^{-1})_{ak}\Big)\ \Big(\sum_m (\boldsymbol{S}^{-1})_{mi}\ \mathbf{Q}_{mj}\Big) \\ &= -\frac{1}{2}(\boldsymbol{S}^{-3}\mathbf{Q})_{ij} - \frac{1}{2}\ \mathrm{tr}(\boldsymbol{S}^{-1})\ (\boldsymbol{S}^{-2}\mathbf{Q})_{ij} - \frac{1}{2}(\boldsymbol{S}^{-3}\mathbf{Q})_{ij} \\ &- \frac{1}{2}\ \mathrm{tr}(\boldsymbol{S}^{-2})\ (\boldsymbol{S}^{-1}\mathbf{Q})_{ij}.\end{aligned}$$

(iv). This is given by the similar way in [4]. □

Proof of Theorem 2.1. The proof is similar to that of Theorem 4.1. in [5]; therefore, we state the outline of the proof only. Let $\hat{\boldsymbol{\Sigma}}_{1k} =$

$\frac{1}{n\gamma(1)}(1+k)\boldsymbol{S},\ |k|<1$ and $\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{n\gamma(1)}\boldsymbol{S}$. We want to show that

$$R_1(\hat{\boldsymbol{\Sigma}}_{1k},\boldsymbol{\Sigma})-R_1(\hat{\boldsymbol{\Sigma}}_1,\boldsymbol{\Sigma}) = E^{(0)}_{\boldsymbol{\Sigma}}\Big[\frac{k}{n\gamma(1)}\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})-p\ \log(1+k)\Big] \geqslant 0 \quad (\forall\boldsymbol{\Sigma}). \tag{6}$$

From Lemma 2.5(i), we can see that

$$R_1(\hat{\boldsymbol{\Sigma}}_{1k},\boldsymbol{\Sigma})-R_1(\hat{\boldsymbol{\Sigma}}_1,\boldsymbol{\Sigma}) = pk - p\log(1+k).$$

Similar to [5], the inequality (6) holds and the proof is complete. □

Proof of Theorem 2.2. The risk of a scalar multiple of $\boldsymbol{S}$ is

$$R_2(a\boldsymbol{S},\boldsymbol{\Sigma}) = E^{(0)}_{\boldsymbol{\Sigma}}\mathrm{tr}(a\boldsymbol{S}\boldsymbol{\Sigma}^{-1}-\mathbf{I})^2 =$$

$$a^2E^{(0)}_{\boldsymbol{\Sigma}}\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) - 2aE^{(0)}_{\boldsymbol{\Sigma}}\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) + pE^{(0)}_{\boldsymbol{\Sigma}}[1].$$

Using (1) and Lemma 2.5 (i) and (ii), we have

$$R_2(a\boldsymbol{S},\boldsymbol{\Sigma}) = \gamma(2)a^2p(n^2+n+np) - 2\gamma(1)nap + p\gamma(0)$$

and the last equality is minimized at $a = \frac{\gamma(1)}{(n+p+1)\gamma(2)}$. □

Proof of Theorem 2.3. Write (2) as

$$\hat{\boldsymbol{\Sigma}} = a\boldsymbol{S} + g(\boldsymbol{S})\mathbf{I} \tag{7}$$

where $g(\boldsymbol{S}) = aut(u)$. Taking differentiating of u with respect to $\boldsymbol{S}$ gives

$$\frac{\partial u}{\partial \boldsymbol{S}} = -(\mathrm{tr}\boldsymbol{S}^{-1})^{-2}\ \frac{\partial(\mathrm{tr}\boldsymbol{S}^{-1})}{\partial \boldsymbol{S}} = u^2\boldsymbol{S}^{-2}.$$

As a result, we obtain

$$\frac{\partial g(\boldsymbol{S})}{\partial \boldsymbol{S}} = a\Big[\frac{\partial u}{\partial \boldsymbol{S}}\ t(u) + u\ \frac{\partial t(u)}{\partial \boldsymbol{S}}\Big] = a\big[u^2t(u) + \frac{\partial t(u)}{\partial u}\ u^3\big]\boldsymbol{S}^{-2}. \tag{8}$$

Let $t(u)$ be a constant function, the equation (8) implies that

$$\frac{\partial g(\boldsymbol{S})}{\partial \boldsymbol{S}} = au^2t\ \boldsymbol{S}^{-2}. \tag{9}$$

From [5] we have the

$$\rho_k = \mathrm{tr}\Big(\frac{\boldsymbol{S}^{-1}}{\mathrm{tr}\boldsymbol{S}^{-1}}\Big)^k, \qquad k = 1, 2, \dots . \tag{10}$$

It is convenient to note that ρ_k decreases in k and $0 \leqslant \rho_k \leqslant 1$.
Use (7) with $a = \frac{1}{n\gamma(1)}$ and

$$\alpha_1 = R_1(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) - R_1(\hat{\boldsymbol{\Sigma}}_1, \boldsymbol{\Sigma}) = E^{(0)}_{\boldsymbol{\Sigma}}\Big[g(\boldsymbol{S})\mathrm{tr}\boldsymbol{\Sigma}^{-1} - \log\big|\mathbf{I} + n\gamma(1)g(\boldsymbol{S})\boldsymbol{S}^{-1}\big|\Big].$$

Knowing that $t(u)$ is a constant, say t, and from (5) with $h = g$, and applying the expansion

$$\log|\mathbf{I} + \alpha\mathbf{B}| = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \; \alpha^n \mathrm{tr}(\mathbf{B}^n)$$

where α is a real number and $\mathbf{B}_{p\times p}$ a symmetric matrix, we obtain

$$\alpha_1 = E^{(1)}_{\boldsymbol{\Sigma}}\Big[(n-p-1)g(\boldsymbol{S})\mathrm{tr}(\boldsymbol{S}^{-1}) + \mathrm{tr}\Big(\frac{\partial g(\boldsymbol{S})}{\partial \boldsymbol{S}}\Big)\Big] - E^{(0)}_{\boldsymbol{\Sigma}}\Big[\sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} \; t^i \rho_i\Big]. \tag{11}$$

A simple calculation shows that

$$\sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} \; t^i \rho_i \geqslant t - \frac{1}{2}t^2,$$

because the terms of the above series are in decreasing magnitude. Substituting (9) into (11), with $a = \frac{1}{n\gamma(1)}$, we obtain

$$\begin{aligned}
\alpha_1 &\leqslant \frac{1}{n\gamma(1)} \; E^{(1)}_{\boldsymbol{\Sigma}}\Big[(n-p-1)t + \rho_2 t\Big] + E^{(0)}_{\boldsymbol{\Sigma}}[\frac{1}{2}t^2 - t], \\
&\leqslant \frac{1}{n\gamma(1)}(n-p)t \; E^{(1)}_{\boldsymbol{\Sigma}}[1] + (\frac{1}{2}t^2 - t) \; E^{(0)}_{\boldsymbol{\Sigma}}[1], \qquad \text{(since } \rho_2 \leqslant 1\text{)}, \\
&\leqslant \big[\frac{n-p}{n} - \gamma(0)\big] \; t + \frac{\gamma(0)}{2} \; t^2 \qquad \text{(from (1))}.
\end{aligned} \tag{12}$$

A sufficient condition for $\alpha_1(\mathbf{\Sigma}) \leqslant 0$ $(\forall\ \Sigma)$ is

$$\big[\frac{n-p}{n} - \gamma(0)\big]\ t + \frac{\gamma(0)}{2}\ t^2 \leqslant 0.$$

Finally, it is seen that

$$0 \leqslant t \leqslant \frac{2\big(n\gamma(0) + p - n\big)}{n\gamma(0)}$$

It is also seen, from (12) that (11) is bounded above by $\big[\frac{n-p}{n} - \gamma(0)\big]\ t + \frac{\gamma(0)}{2}\ t^2$; hence, the optimal t is $\frac{n\gamma(0)+p-n}{n\gamma(0)}$. □

Theorem 2.4. compares $\hat{\mathbf{\Sigma}}$ in (7) and $\hat{\mathbf{\Sigma}}_2 = a\boldsymbol{S}$, $a = \frac{\gamma(1)}{(n+p+1)\gamma(2)}$. Let

$$\begin{aligned}\alpha_2(\mathbf{\Sigma}) &= R_2(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) - R_2(\hat{\mathbf{\Sigma}}_2, \mathbf{\Sigma})\\ &= E^{(0)}_{\mathbf{\Sigma}}\Big[2a\ g(\boldsymbol{S})\mathrm{tr}(\boldsymbol{S}\mathbf{\Sigma}^{-2})) - 2\ g(\boldsymbol{S})\mathrm{tr}(\Sigma^{-1} + g^2(\boldsymbol{S})\mathrm{tr}(\mathbf{\Sigma}^{-2})\Big]\end{aligned} \tag{13}$$

Then identities for the terms in (13) are given in the following Lemma.

Lemma 2.7. *Under conditions of Lemma* 5.2.[5]*(see appendix)*

i) *For* $h(\boldsymbol{S}) = 2a\ g(\boldsymbol{S})$, *we have*

$$\begin{aligned}E^{(0)}_{\mathbf{\Sigma}}[h(\boldsymbol{S})\ \mathrm{tr}(\mathbf{\Sigma}^{-2}\boldsymbol{S})] =& E^{(2)}_{\mathbf{\Sigma}}\Big[n(n-p-1)h(\boldsymbol{S})\ \mathrm{tr}(\boldsymbol{S}^{-1})\\ &- (p+1)\mathrm{tr}\big(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\big)\\ &+ 2\mathrm{tr}D_{\boldsymbol{S}}\big(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}} \cdot \boldsymbol{S}\big)\Big]\end{aligned}$$

ii) *For* $h(\boldsymbol{S}) = -2g(\boldsymbol{S})$, *we obtain*

$$E^{(0)}_{\mathbf{\Sigma}}[h(\boldsymbol{S})\ \mathrm{tr}(\mathbf{\Sigma}^{-1})] = E^{(1)}_{\mathbf{\Sigma}}\Big[(n-p-1)h(\boldsymbol{S})\ \mathrm{tr}(\boldsymbol{S}^{-1}) + \mathrm{tr}\big(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\big)\Big]$$

iii) *For* $h(\boldsymbol{S}) = g^2(\boldsymbol{S})$, *we get*

$$E^{(0)}_{\boldsymbol{\Sigma}}[h(\boldsymbol{S})\ \mathrm{tr}(\boldsymbol{\Sigma}^{-2})] = E^{(2)}_{\boldsymbol{\Sigma}}\Big[2\ \mathrm{tr} D_S\big(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\big) + 2(n-p-1)\ \mathrm{tr}\big(\boldsymbol{S}^{-1}\ \frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\big) + (n-p-1)\ (n-p-2)h(\boldsymbol{S})\ \mathrm{tr}(\boldsymbol{S}^{-2}) - (n-p-1)h(\boldsymbol{S})\ \mathrm{tr}^2(\boldsymbol{S}^{-1})\Big]$$

Proof (i). Set $\mathbf{T} = \boldsymbol{S}\boldsymbol{\Sigma}^{-1}h(\boldsymbol{S})$, by using (4), we can see

$$E^{(0)}_{\boldsymbol{\Sigma}}\Big[\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}h(\boldsymbol{S})\big)\Big] = E^{(1)}_{\Sigma}\Big[(n-p-1)\ \mathrm{tr}\big(h(\boldsymbol{S})\ \boldsymbol{\Sigma}^{-1}\big) + 2\ \mathrm{tr} D_{\boldsymbol{S}}\big(h(\boldsymbol{S})\ \boldsymbol{S}\boldsymbol{\Sigma}^{-1}\big)\Big].$$

Now applying Lemma 2.6 (i) and (ii) with $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$ to the second term $D_{\boldsymbol{S}}\big(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}h(\boldsymbol{S})\big)$; hence, we get

$$E^{(0)}_{\boldsymbol{\Sigma}}\Big[h(\boldsymbol{S})\ \mathrm{tr}(\boldsymbol{\Sigma}^{-2}\boldsymbol{S})\Big] = E^{(1)}_{\boldsymbol{\Sigma}}\Big[n\ \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}h(\boldsymbol{S})\big) + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\ \frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\boldsymbol{S}\big)\Big].$$

The result obtains by applying (5) to the first term and then (4) to the second.

(ii). In (5), set $k = 0$.

(iii). Put $\mathbf{T} = \boldsymbol{\Sigma}^{-1}h(\boldsymbol{S})$. From (4) and Lemma 2.6. (i), we have

$$E^{(0)}_{\boldsymbol{\Sigma}}\Big[\mathrm{tr}\big(\boldsymbol{\Sigma}^{-2}h(\boldsymbol{S})\big)\Big] = E^{(1)}_{\boldsymbol{\Sigma}}\Big[(n-p-1)\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}^{-1}h(\boldsymbol{S})\big) + \mathrm{tr}\big(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\boldsymbol{\Sigma}^{-1}\big)\Big].$$

Applying (4) to the first term with $\mathbf{T} = \boldsymbol{S}^{-1}h(\boldsymbol{S})$ and applying it to the second with $\mathbf{T} = \frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}$, we obtain

$$E^{(0)}_{\boldsymbol{\Sigma}}\Big[\mathrm{tr}\big(\boldsymbol{\Sigma}^{-2}h(\boldsymbol{S})\big)\Big] = E^{(2)}_{\boldsymbol{\Sigma}}\Big[(n-p-1)^2h(\boldsymbol{S})\mathrm{tr}(\boldsymbol{S}^{-2}) + 2(n-p-1)\ \mathrm{tr} D_{\boldsymbol{S}}(\boldsymbol{S}^{-1}h(\boldsymbol{S})) + (n-p-1)\mathrm{tr}\big(\boldsymbol{S}^{-1}\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\big)\ 2\mathrm{tr} D_{\boldsymbol{S}}\big(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\big)\Big].$$

Our result yields by using Lemma 2.6. (i) with $\mathbf{F} = \boldsymbol{S}^{-1}$ and $\phi = h$ and Lemma 2.6. (iv) with $\mathbf{Q} = \mathbf{I}$ to the second term.
The special of Lemma 2.7. by taking $g(\boldsymbol{S}) = aut$, $u = (\mathrm{tr}(\boldsymbol{S}^{-1})^{-1}$ and $a = \frac{\gamma(1)}{(n+p+1)\gamma(2)}$ is the following lemma. □

Lemma 2.8. *For $g(S) = aut$,*

i) $E^{(0)}_{\boldsymbol{\Sigma}}\Big[2a^2ut\ \mathrm{tr}(\boldsymbol{\Sigma}^{-2}\boldsymbol{S})\Big] = E^{(2)}_{\boldsymbol{\Sigma}}\Big[2na^2(n-p-1)t - 2a^2\rho_2(p+2)t + 4a^2t\rho_3 - 2a^2t\Big],$

ii) $E^{(0)}_{\boldsymbol{\Sigma}}\Big[-2aut\ \mathrm{tr}(\boldsymbol{\Sigma}^{-1})\Big] = E^{(1)}_{\boldsymbol{\Sigma}}\Big[-2a(n-p-1)t - 2at\rho_2\Big],$

iii) $E^{(0)}_{\boldsymbol{\Sigma}}\Big[a^2u^2t^2\ \mathrm{tr}(\boldsymbol{\Sigma}^{-2})\Big] = E^{(2)}_{\boldsymbol{\Sigma}}\Big\{-(n-p-1)a^2t^2 + 6a^2\rho_4t^2 + 4a^2t^2\rho_3(n-p-2) + a^2t\rho_2[(n-p-1)^2t - (n-p-1)t - 4t]\Big\}$

Proof (i). In Lemma 2.7. (i), we put $h(\boldsymbol{S}) = 2a^2ut$ and then compute $D_{\boldsymbol{S}}(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\boldsymbol{S})$. By applying Lemma 2.6. (i), (iv) with $\phi = u^2$, $\mathbf{F} = \boldsymbol{S}^{-1}$ and $\mathbf{Q} = \mathbf{I}$, we have

$$\begin{aligned}
\mathrm{tr}D_{\boldsymbol{S}}\big(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\cdot\boldsymbol{S}\big) &= \mathrm{tr}\Big\{2a^2t\ \Big[\frac{1}{2}\frac{\partial u^2}{\partial \boldsymbol{S}}\boldsymbol{S}^{-1} + u^2\big\{-\frac{1}{2}\boldsymbol{S}^{-2} - \frac{1}{2}\mathrm{tr}(\boldsymbol{S}^{-1})\boldsymbol{S}^{-1}\big\}\Big]\Big\}\\
&= \mathrm{tr}\Big\{2a^2t\ u^3\boldsymbol{S}^{-3} - a^2t\ u^2\boldsymbol{S}^{-2} - a^2t\ u^2\mathrm{tr}(\boldsymbol{S}^{-1})\boldsymbol{S}^{-1}\Big\}\\
&= 2a^2t\rho_3 - a^2t\rho_2 - a^2t,
\end{aligned}$$

Result (i) is now obtainable by using (9).

(ii). Using (9) and (10), the result obtains from Lemma 2.7. (ii) with $h(\boldsymbol{S}) = -2aut$.

(iii). Take $h(\boldsymbol{S}) = a^2u^2t^2$ in Lemma 2.7. (iii); then compute $D_{\boldsymbol{S}}(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}})$. Similarly, using Lemma 2.6. (i), (iii) with $\phi = u^3$, $\mathbf{F} = \boldsymbol{S}^{-2}$

and $\mathbf{Q} = \mathbf{I}$, we have

$$\begin{aligned}\mathrm{tr} D_{\boldsymbol{S}}\Big(\frac{\partial h(\boldsymbol{S})}{\partial \boldsymbol{S}}\Big) &= 2a^2t^2 \ \mathrm{tr}\Big[\frac{1}{2}\Big(\frac{\partial u^3}{\partial \boldsymbol{S}}\boldsymbol{S}^{-2}\Big) + u^3 D_{\boldsymbol{S}}\boldsymbol{S}^{-2}\Big] \\ &= a^2t^2[3\rho_4 - 2\rho_3 - 2\rho_2]. \end{aligned} \tag{14}$$

Now substitute (14) in Lemma 2.7. (iii) to achieve the result. □

Proof of Theorem 2.4. Each term in (13) obtained from Lemma 2.7 (i), (ii) and (iii); hence, we have

$$\begin{aligned}\alpha_2 = E^{(2)}_{\boldsymbol{\Sigma}}\Big[&a^2t\big\{2n(n-p-1) + 2\rho_2(p+2) + 4\rho_3 - 2\big\} \\ &+ a^2t^2\big\{-(n-p-1) + 6\rho_4 + 4\rho_3(n-p-2) + \rho_2(n-p-1)^2 \\ &- \rho_2(n-p-1) - 4\rho_2\big\}\Big] + E^{(1)}_{\boldsymbol{\Sigma}}\Big[at\big\{-2(n-p-1) - 2\rho_2\big\}\Big].\end{aligned} \tag{15}$$

The above coefficient of a^2t can be written as $2[n(n-p-1)-1] + 2\rho_2(p+2) + 4\rho_3$. Since $\rho_k \leqslant 1$ thus, the entire quantity is bounded above by

$$2[n(n-p-1)-1] + 2(p+2) + 4 = 2[(n-1)(n-p)+3]. \tag{16}$$

Since $\rho_k \searrow$, $0 \leqslant \rho_k \leqslant 1$ and $n > p+1$, the coefficient of a^2t^2 is bounded above by

$$\rho_2[\,2(n-p-1) + (n-p-1)^2 - 2]$$

and the above term is bounded by

$$\rho_2[\,2(n-p-1) + (n-p-1)^2].$$

The above coefficient of ρ_2 is positive and $\rho_2 \leqslant 1$; accordingly, the coefficient of a^2t^2 is bounded above by

$$(n-p-1)(n-p+1). \tag{17}$$

From [1], paper 137, we have $\frac{1}{p} \leqslant \rho_2 \leqslant 1$ then in (15), for the coefficient of at we can see

$$-2(n-p-1) - 2\rho_2 \leqslant -\frac{2}{p}. \tag{18}$$

Finally, from (15)-(18) a sufficient condition for $\alpha_2(\Sigma) \leqslant 0$ ($\forall\ \Sigma$) is

$$2a^2t\big[(n-1)(n-p)+3\big]\gamma(2)+a^2t^2(n-p-1)(n-p+1)\gamma(2)+at(-\frac{2}{p})\gamma(1) \leqslant 0. \tag{19}$$

Recall that $a = \frac{\gamma(1)}{\gamma(2)(n+p+1)}$ then (19) is equivalent to

$$0 \leqslant t \leqslant \frac{2\big[(n-2p+1)-p(n-1)(n-p)\big]}{p(n-p-1)(n-p+1)}.$$

The left hand side of (19) is minimized at $t = \frac{(n-2p+1)-p(n-1)(n-p)}{p(n-p-1)(n-p+1)}$. □

Appendix

Lemma 2.1. *Let* $\mathbf{Q}$ *be a* $p \times p$ *matrix of constants. If*

$$\lim_{z\to\pm\infty} |z| f^{(1)}(z^2+a^2) = 0$$

for any real a*, then we have*

(i) $\mathbb{E}_\Sigma^{(0)}[\mathbf{S}] = n\gamma(1)\mathbf{\Sigma}$.

If $\lim_{z\to\pm\infty} |z|^3 f^{(1)}(z^2+a^2) = 0$ *and* $\lim_{z\to\pm\infty} |z| f^{(2)}(z^2+a^2) = 0$

for any real a*, thenwe have*

(ii) $\mathbb{E}_\Sigma^{(0)}[\mathbf{SQS}] = \gamma(2)(n^2\mathbf{\Sigma Q \Sigma} + n\mathbf{\Sigma \acute{Q} \Sigma} + ntr(\mathbf{Q\Sigma})\mathbf{\Sigma})$,

(iii) $\mathbb{E}_\Sigma^{(0)}[tr(\mathbf{QS})\mathbf{S}] = \gamma(2)(n^2 tr(\mathbf{Q\Sigma})\mathbf{\Sigma} + n\mathbf{\Sigma Q \Sigma} + n\mathbf{\Sigma \acute{Q} \Sigma})$.

Lemma 5.2. *Let* $\mathbf{G}(\mathbf{S}) = (g_{ab}(\mathbf{S})) = \mathbf{G}(\sum_{\mathbf{c=1}}^{\mathbf{n}} \mathbf{\acute{z}_c z_c})$*, be a* $p \times p$ *matrix whose elements are differentiable with respect to* $z_{jk}(j = 1,2,...,n, k = 1,2,...,p)$*. For* $-2 \leqslant i \leqslant 1$*, assume that*

(a) $\mathbb{E}_\Sigma^i[|g_{ab}(\mathbf{S}))|] < \infty$;

(b) $\lim_{zjk\to\pm\infty} |z_{jk}| \mathbf{G}(\sum_{\mathbf{c=1}}^{\mathbf{n}} \mathbf{\acute{z}_c z_c})(\sum_{\mathbf{c=1}}^{\mathbf{n}} \mathbf{\acute{z}_c z_c})^{-1} f^{i+1}(z_{jk}^2 + a^2) = \mathbf{0}_{p\times p}$ *for any real* a*.*

Then we have

$$\mathbb{E}_\Sigma^{(i)}[\mathbf{\Sigma^{-1}G(S)}] = \mathbb{E}_\Sigma^{(i+1)}[(n-p-1)\mathbf{S^{-1}G(S)} + \mathbf{2D_s G(S)}].$$

References

[1] R. Bellman, *Introduction to Matrix Analysis*, New York: McGraw-Hill, 1970.

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

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

[4] L. R. Haff, Identities for the inverse Wishart distribution with computational results in linear and quadradic discrimination, *Sankhaya, B*, 44 (1982), 245-258.

[5] T. Hisayuki, Estimating the inverse matrix of scale parameter in an elliptically contoured distribution, *J. Japan Statist. Soc.*, 35 (2005), 21-39.

[6] W. James and C. Stein, Estimation with quadratic loss. In *Fourth Berkeley Symp. Math. Statist. Probability* Univ. california Press, Berkeley, 1961.

[7] T. Kubokawa and M. S. Srivastava, Roubst improvement in estimation of a covariance matrix in an elliptically contoured distribution, *Ann. Statist.*, 27 (1999), 600-609.

[8] M. J. Daniels and M. Pourahmadi, Bayesian analysis of covariance matrices and dynamic models for longitudinal data,*Biometrika.*, 89 (2002), 553-566.

[9] M. D. Perlman, Reduced mean square estimation for several parameters. *Sankhaya, B*, 34 (1972), 89-92.

[10] J. G. Scott and J. O. Berger, Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem, *Ann. Statist.*, 38 (2010), 2587-2619.

[11] J. Selliah, *Estimation and Testing Problems in a Wishart Distribution,* Ph. D. thesis, Department of Statistics, Standford University, 1964.

Zahra Khodadadi
Department of Mathematics
Assistant Professor of Mathematics
Islamic Azad University-Marvdasht Branch
Marvdasht, Iran.
E-mail: zhr_ khodadady@yahoo.com

Bahram Tarami
Department of Mathematics
Assistant Professor of Mathematics
College of Sciences, Yasouj University
Yasouj, Iran.
E-mail: tarami@mail.yu.ac.ir