The Annals of Statistics
1999, Vol. 27, No. 2, 600–609

ROBUST IMPROVEMENT IN ESTIMATION OF A COVARIANCE MATRIX IN AN ELLIPTICALLY CONTOURED DISTRIBUTION

BY T. KUBOKAWA[1] AND M. S. SRIVASTAVA[2]

University of Tokyo and University of Toronto

This paper derives an extended version of the Haff or, more appropriately, Stein–Haff identity for an elliptically contoured distribution (ECD). This identity is then used to show that the minimax estimators of the covariance matrix obtained under normal models remain robust under the ECD model.

1. Introduction. Consider the multivariate linear regression model

$$\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \mathbf{e} \tag{1.1}$$

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

$$|\boldsymbol{\Sigma}|^{-N/2} f(\mathrm{tr}\, \boldsymbol{\Sigma}^{-1}\mathbf{e}^t\mathbf{e}), \tag{1.2}$$

where $\boldsymbol{\Sigma}$ is a $p \times p$ unknown positive-definite matrix, $f(\cdot)$ is a nonnegative unknown function on the nonnegative real line, $\mathbf{e}^t$ denotes the transpose of the matrix $\mathbf{e}$ and $\mathrm{tr}(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$. The model (1.2) is called the *elliptically contoured distribution* (ECD) which we shall refer to as the ECD model in this paper. It may be noted that the density $f(\cdot)$ depends on N, but for simplicity of notation this dependence is not shown.

Beginning with the seminal works of Stein (1956) and James and Stein (1961), the problem of estimating the matrix of regression parameters $\boldsymbol{\beta}$ under a squared loss function has been considered many times in statistical literature for the normal model. See, for example, Robert (1994) and Kubokawa (1998). Robustness of these procedures under the ECD model, however, has been considered only in the last decade. For example, Srivastava and Bilodeau (1989) established the robustness of the Stein estimator when the error matrix has the distribution of a scale mixture (with signed

Received August 1997; revised December 1998.

[1]Supported in part by the Shirakawa Institute of Animal Genetics and the Japan Livestock Technology Association, by the Ministry of Education, Japan, Grants 08780216, 09780214 and by a grant from the Research Institute for the Japanese Economy, University of Tokyo.

[2]Supported by Natural Sciences and Engineering Research Council of Canada.

AMS 1991 *subject classifications*. Primary 62H12; secondary 62F11.

Key words and phrases. Elliptically contoured distribution, robustness of improvement, multivariate linear model, covariance matrix, statistical decision theory, shrinkage estimation.

measure) of multivariate normal distributions and Cellier, Fourdrinier and Robert (1989) showed the robustness for $p = 1$ by extending Stein's identity to the ECD model. Most recently Kubokawa and Srivastava (1997) made no such simplifying assumption and showed that the minimax estimators obtained by Bilodeau and Kariya (1989) under the normal model remain robust under the ECD model.

The problem of estimating the covariance matrix $\mathbf{\Sigma}$ under Stein's loss has also been considered in the literature and minimax estimators were obtained by James and Stein (1961) and Dey and Srinivasan (1985) for the normal model. It is not known, however, whether these minimax estimators remain robust under the ECD model.

In this paper, we consider the problem of estimating the scale matrix $\mathbf{\Sigma}$ for the ECD model in a decision-theoretic set-up. The performance of every estimator is evaluated in terms of a risk function. Most results in the normal model employ the integration by parts approach in the Wishart distribution derived by Stein (1977a) and Haff (1979), known in the literature as the "Haff identity." We first extend this identity to the ECD model. Since our approach and proofs are based on Stein's method, we shall more appropriately call it the Stein–Haff identity for the ECD model. Using this extended identity, we establish that the dominance results given by James and Stein (1961) and Dey and Srinivasan (1985) remain robust under the ECD model.

2. Main results. Let $\mathbf{S}$ be a random matrix having a Wishart distribution with n degrees of freedom and mean $E[\mathbf{S}] = n\mathbf{\Sigma}$. We shall consider the problem of estimating $\mathbf{\Sigma}$ by $\hat{\mathbf{\Sigma}}$ that minimizes the risk for the Stein loss function given by $\operatorname{tr} \hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1} - \log|\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}| - p$. If we restrict our attention to estimators of the kind $a\mathbf{S}$ where a is a scalar, then the unbiased estimator $\hat{\mathbf{\Sigma}}^{UB} = n^{-1}\mathbf{S}$ is the best estimator in the sense that it has the minimum risk for Stein's loss defined above. It was, however pointed out by Stein (1975) that the eigenvalues of $\hat{\mathbf{\Sigma}}^{UB}$ spread out more than the corresponding eigenvalues of $\mathbf{\Sigma}$. This problem is more serious when $\mathbf{\Sigma} \cong \mathbf{I}_p$. This fact suggests that $\hat{\mathbf{\Sigma}}^{UB}$ should be shrunk toward a middle value. This phenomenon is similar to the Stein-type estimation of a multivariate normal mean vector [see Stein (1975, 1977a, b), Yang and Berger (1994) and the references therein].

Initially, James and Stein (1961) considered the problem of obtaining minimax estimators of $\mathbf{\Sigma}$. By considering the best equivariant estimator with respect to the triangular group, they obtained a minimax estimator of the form

$$\hat{\mathbf{\Sigma}}^{JS} = \mathbf{T}\mathbf{D}\mathbf{T}^t,$$

where $\mathbf{D} = \operatorname{diag}(d_1, \dots, d_p)$ with $d_i = (n + p + 1 - 2i)^{-1}$, $i = 1, \dots, p$ and $\mathbf{T}$ is a $p \times p$ lower triangular matrix with positive diagonal elements such that $\mathbf{S} = \mathbf{T}\mathbf{T}^t$. They showed that $\hat{\mathbf{\Sigma}}^{JS}$ has smaller risk than $\hat{\mathbf{\Sigma}}^{UB}$ for Stein's loss. The estimator $\hat{\mathbf{\Sigma}}^{JS}$ has, however, the drawback that it depends on the coordinate system. Thus it will be desirable to construct orthogonally invariant minimax estimators. Stein (1977a, b) and Dey and Srinivasan (1985) obtained an orthogonally invariant estimator

$$\hat{\mathbf{\Sigma}}^{SDS} = \mathbf{H}\,\mathrm{diag}(d_1 l_1, \dots, d_p l_p)\mathbf{H}^t,$$

where $\mathbf{H}$ is a $p \times p$ orthogonal matrix and $l_1, \dots, l_p$ are the ordered eigenvalues of the random matrix $\mathbf{S}$ such that $\mathbf{S} = \mathbf{H}\,\mathrm{diag}(l_1, \dots, l_p)\mathbf{H}^t$. They showed that the estimator $\hat{\mathbf{\Sigma}}^{SDS}$ dominates $\hat{\mathbf{\Sigma}}^{JS}$ for Stein's loss. On the other hand, Takemura (1984) gave an orthogonally invariant estimator of the form

$$\hat{\mathbf{\Sigma}}^{TK} = \int_{O(p)} \mathbf{\Gamma}\mathbf{T}_\Gamma \mathbf{D}\mathbf{T}_\Gamma^t \mathbf{\Gamma}^t \, d\mathbf{\Gamma},$$

where $O(p)$ denotes a class of $p \times p$ orthogonal matrices and $\mathbf{T}_\Gamma$ is a $p \times p$ lower triangular matrix with positive diagonal element such that $\mathbf{\Gamma}^t\mathbf{S}\mathbf{\Gamma} = \mathbf{T}_\Gamma\mathbf{T}_\Gamma^t$. This estimator was also shown to dominate $\hat{\mathbf{\Sigma}}^{JS}$. The orthogonally invariant minimax estimator $\hat{\mathbf{\Sigma}}^{SDS}$, however, has another problem: the diagonal elements $d_1 l_1, \dots, d_p l_p$ are not ordered. It would be desirable to have the ordered values as diagonal elements if the dominance results still hold. Recently, Sheena and Takemura (1992) considered estimators of the above kind but with ordered diagonal elements and showed that these estimators dominate the unordered estimators. In other words, let $\hat{\mathbf{\Sigma}}(\boldsymbol{\phi}) = \mathbf{H}\,\mathrm{diag}(\phi_1(\mathbf{L}), \dots, \phi_p(\mathbf{L}))\mathbf{H}^t$ be an orthogonally invariant estimator and let $\hat{\mathbf{\Sigma}}(\boldsymbol{\phi}^O)$ be the order-preserving estimator given by modifying $\hat{\mathbf{\Sigma}}(\boldsymbol{\phi})$ as $\hat{\mathbf{\Sigma}}(\boldsymbol{\phi}^O) = \mathbf{H}\,\mathrm{diag}(\phi_1^O(\mathbf{L}), \dots, \phi_p^O(\mathbf{L}))\mathbf{H}^t$, where $\phi_i^O(\mathbf{L})$ is the ith largest element in $(\phi_1(\mathbf{L}), \dots, \phi_p(\mathbf{L}))$, that is, $\phi_1^O(\mathbf{L}) \ge \cdots \ge \phi_p^O(\mathbf{L})$. Then $\hat{\mathbf{\Sigma}}(\boldsymbol{\phi}^O)$ is better than $\hat{\mathbf{\Sigma}}(\boldsymbol{\phi})$ in the normal distribution if $P_\Sigma[\phi_i^O(\mathbf{L}) \ne \phi_i(\mathbf{L})$ for some $i] > 0$ for some $\mathbf{\Sigma}$. This result can be applied to the nonorder-preserving estimator $\hat{\mathbf{\Sigma}}^{SDS}$, which demonstrates the inadmissibility of $\hat{\mathbf{\Sigma}}^{SDS}$. For numerical comparison of the above-mentioned estimators for $p = 2$, see Sugiura and Ishibayashi (1997) who also showed that the reference prior Bayes estimator given by Yang and Berger (1994) is superior to $\hat{\mathbf{\Sigma}}^{SDS}$ and $\hat{\mathbf{\Sigma}}^{TK}$ when $\mathbf{\Sigma} \cong \mathbf{I}_p$ for $n \ge 3$ although it has a risk slightly larger than $\hat{\mathbf{\Sigma}}^{JS}$ when $\mathbf{\Sigma}$ is far from $\mathbf{I}_p$.

Our objective is to establish that the above dominance results hold for every ECD model, that is, the improvement is robust. For the purpose, we first derive an extended version of the Stein-Haff identity for the ECD model.

Let $\mathbf{P}$ be an $N \times N$ orthogonal matrix such that $(\mathbf{PA})^t = ((\mathbf{A}^t\mathbf{A})^{1/2}, \mathbf{0})$ and let $\boldsymbol{\theta} = (\mathbf{A}^t\mathbf{A})^{1/2}\boldsymbol{\beta}$. Let $\mathbf{x}$ and $\mathbf{z}$ be, respectively, $m \times p$ and $n \times p$ matrices such that $(\mathbf{x}^t, \mathbf{z}^t)^t = \mathbf{Py}$ for $n = N - m$; then the joint density of $\mathbf{x}$ and $\mathbf{z}$ has the form

$$|\mathbf{\Sigma}|^{-N/2} f\big(\mathrm{tr}\,\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\theta})^t(\mathbf{x} - \boldsymbol{\theta}) + \mathrm{tr}\,\mathbf{\Sigma}^{-1}\mathbf{z}^t\mathbf{z}\big). \tag{2.1}$$

Denote $\mathbf{S} = \mathbf{z}^t\mathbf{z}$. We treat the estimation issue of $\mathbf{\Sigma}$ based on $\mathbf{x}$ and $\mathbf{S}$. Let

$F(x) = 2^{-1}\int_x^{+\infty} f(t)\,dt$ and define

$$E_{\theta,\Sigma}^{f} h(\mathbf{x},\mathbf{z}) = \int\int h(\mathbf{x},\mathbf{z})|\mathbf{\Sigma}|^{-N/2} f\big(\operatorname{tr}\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\theta})^t(\mathbf{x}-\boldsymbol{\theta}) + \operatorname{tr}\mathbf{\Sigma}^{-1}\mathbf{z}^t\mathbf{z}\big)\,d\mathbf{x}\,d\mathbf{z},$$

$$E_{\theta,\Sigma}^{F} h(\mathbf{x},\mathbf{z}) = \int\int h(\mathbf{x},\mathbf{z})|\mathbf{\Sigma}|^{-N/2} F\big(\operatorname{tr}\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\theta})^t(\mathbf{x}-\boldsymbol{\theta}) + \operatorname{tr}\mathbf{\Sigma}^{-1}\mathbf{z}^t\mathbf{z}\big)\,d\mathbf{x}\,d\mathbf{z},$$

where $h(\mathbf{x},\mathbf{z})$ is an integrable function. When there is no confusion, we shall drop $\boldsymbol{\theta}$ from the subscript in the above definitions. Let $\mathbf{G}(\mathbf{S})$ be a $p \times p$ matrix such that the (i,j) element $g_{ij}(\mathbf{S})$ is a function of $\mathbf{S} = (s_{ij})$ and denote

$$\{\mathbf{D}_S\mathbf{G}(\mathbf{S})\}_{ij} = \sum_a d_{ia} g_{aj}(\mathbf{S}),$$

where

$$d_{ia} = \frac{1}{2}(1+\delta_{ia})\frac{\partial}{\partial s_{ia}},$$

with $\delta_{ia} = 1$ for $i = a$ and $\delta_{ia} = 0$ for $i \neq a$. Note that $\mathbf{S} = \sum_{i=1}^n \mathbf{z}_i^t\mathbf{z}_i$ for $n = N - m$, $\mathbf{z} = (\mathbf{z}_1^t,\dots,\mathbf{z}_n^t)^t$ and $\mathbf{z}_k = (z_{k1},\dots,z_{kp})$. Then we get the following extended version of the Stein–Haff identity.

Lemma 1. *For $k = 1,\dots,n$ and $j = 1,\dots,p$, assume that $\mathbf{G}(\sum_{i=1}^n \mathbf{z}_i^t\mathbf{z}_i)$ is differentiable with respect to z_{kj} and that*

(a) $E_{\theta,\Sigma}^{f}[|\operatorname{tr}\{\mathbf{G}(\mathbf{S})\mathbf{\Sigma}^{-1}|]$ *is finite*;

(b) $\lim_{z_{kj}\to\pm\infty} |z_{kj}| \mathbf{G}(\sum_{i=1}^n \mathbf{z}_i^t\mathbf{z}_i)(\sum_{i=1}^n \mathbf{z}_i^t\mathbf{z}_i)^{-1} F(z_{kj}^2 + a^2) = \mathbf{0}$ *for any real a. Then*

$$E_{\theta,\Sigma}^{f}\big[\operatorname{tr}\{\mathbf{G}(\mathbf{S})\mathbf{\Sigma}^{-1}\}\big] = E_{\theta,\Sigma}^{F}\big[(n-p-1)\operatorname{tr}\{\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}\} + 2\operatorname{tr}\{\mathbf{D}_S\mathbf{G}(\mathbf{S})\}\big].$$

The proof of this lemma is deferred to the appendix. Based on Lemma 1, we prove the robustness of the two dominance results: $\hat{\mathbf{\Sigma}}^{JS}$ improving $\hat{\mathbf{\Sigma}}^{UB}$ and $\hat{\mathbf{\Sigma}}^{SDS}$ improving $\hat{\mathbf{\Sigma}}^{JS}$.

Proposition 1. *For the estimation of $\mathbf{\Sigma}$ in the canonical form* (2.1), *the James–Stein estimator $\hat{\mathbf{\Sigma}}^{JS}$ is better than $\hat{\mathbf{\Sigma}}^{UB}$ uniformly for every unknown function $f(\cdot)$. Also the orthogonally invariant estimator $\hat{\mathbf{\Sigma}}^{TK}$ is superior to $\hat{\mathbf{\Sigma}}^{JS}$ uniformly for every unknown function $f(\cdot)$.*

Proof. The risk difference of the estimators $\hat{\mathbf{\Sigma}}^{UB}$ and $\hat{\mathbf{\Sigma}}^{JS}$ relative to Stein's loss is written as

$$\begin{aligned}\Delta_1 &= R\big(\hat{\mathbf{\Sigma}}^{UB},(\boldsymbol{\theta},\mathbf{\Sigma}),f\big) - R\big(\hat{\mathbf{\Sigma}}^{JS},(\boldsymbol{\theta},\mathbf{\Sigma}),f\big)\\ &= E_{\Sigma}^{f}\big[n^{-1}\operatorname{tr}\mathbf{S}\mathbf{\Sigma}^{-1} - \log|n^{-1}\mathbf{S}\mathbf{\Sigma}^{-1}| - \operatorname{tr}\mathbf{TDT}^t\mathbf{\Sigma}^{-1} + \log|\mathbf{TDT}^t\mathbf{\Sigma}^{-1}|\big]\\ &= E_{I}^{f}\left[n^{-1}\operatorname{tr}\mathbf{S} + p\log n - \operatorname{tr}\mathbf{TDT}^t + \sum_{i=1}^{p}\log d_i\right].\end{aligned} \tag{2.2}$$

From Lemma 1, we have

$$E_I^f[n^{-1}\operatorname{tr}\mathbf{S}] = E_I^F\left[(n-p-1)\frac{1}{n}\operatorname{tr}\mathbf{I}_p + \frac{2}{n}\frac{p+1}{2}\operatorname{tr}\mathbf{I}_p\right] \tag{2.3}$$
$$= p \times E_I^F[1].$$

If we can show that

$$E_I^f[\mathbf{T}^t\mathbf{T}] = \mathbf{D}^{-1}E_I^F[1], \tag{2.4}$$

then combining (2.2), (2.3) and (2.4) gives

$$\Delta_1 = E_I^f\left[p\log n - \sum_{i=1}^{p}\log(n+p+1-2i)\right],$$

which is nonnegative, as checked easily.

We shall verify the condition (2.4) to complete the proof. For the purpose, $\mathbf{S}$ and $\mathbf{T}$ are decomposed by $\mathbf{S} = (\mathbf{S}_{ij})$ and $\mathbf{T} = (\mathbf{T}_{ij})$ for $i, j = 1, 2$ with scalars S_{22}, T_{22} and $\mathbf{T}_{12} = \mathbf{0}$. Since $\mathbf{S}_{11} = \mathbf{T}_{11}\mathbf{T}_{11}^t$, $\mathbf{S}_{12} = \mathbf{T}_{11}\mathbf{T}_{21}^t$ and $S_{22} = \mathbf{T}_{21}\mathbf{T}_{21}^t + T_{22}^2$, we observe that

$$(\mathbf{T}^t\mathbf{T})_{11} = \mathbf{T}_{11}^t\mathbf{T}_{11} + \mathbf{T}_{21}^t\mathbf{T}_{21} = \mathbf{T}_{11}^t\mathbf{T}_{11} + \mathbf{T}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{12}^t\left(\mathbf{T}_{11}^t\right)^{-1},$$
$$(\mathbf{T}^t\mathbf{T})_{12} = \mathbf{T}_{21}^t T_{22} = \mathbf{T}_{11}^{-1}\mathbf{S}_{12}\sqrt{S_{22} - \mathbf{S}_{12}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}},$$
$$(\mathbf{T}^t\mathbf{T})_{22} = T_{22}^2 = S_{22} - \mathbf{S}_{12}^t\mathbf{S}_{11}^{-1}\mathbf{S}_{12}.$$

Let $\mathbf{S}_{ij} = \mathbf{u}_i^t\mathbf{u}_j$ for $\mathbf{z} = (\mathbf{u}_1, \mathbf{u}_2)$ with $n \times 1$ vector $\mathbf{u}_2$, and let $(\mathbf{v}_1^t, \mathbf{v}_2^t)^t = \mathbf{Q}\mathbf{u}_2$ with $(p-1)\times 1$ vector $\mathbf{v}_1$ for $n \times n$ orthogonal matrix $\mathbf{Q}$ such that $(\mathbf{Q}\mathbf{u}_1)^t = (\mathbf{T}_{11}, \mathbf{0})$. Then the joint density of $(\mathbf{x}, \mathbf{u}_1, \mathbf{v}_1, \mathbf{v}_2)$ is written by

$$|\boldsymbol{\Sigma}|^{-m/2} f\left(\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\theta})^t(\mathbf{x}-\boldsymbol{\theta}) + \operatorname{tr}\mathbf{u}_1^t\mathbf{u}_1 + \mathbf{v}_1^t\mathbf{v}_1 + \mathbf{v}_2^t\mathbf{v}_2\right).$$

Since $\mathbf{S}_{12} = \mathbf{u}_1^t\mathbf{u}_2 = \mathbf{T}_{11}\mathbf{v}_1$, the same argument used in (A.1) of the Appendix gives that

$$E_I^f\left[\mathbf{T}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{12}^t\left(\mathbf{T}_{11}^t\right)^{-1}\right] = E_I^f\left[\mathbf{v}_1\mathbf{v}_1^t\right] = \mathbf{I}_{p-1}E_I^F[1], \tag{2.5}$$

$$E_I^f\left[S_{22} - \mathbf{S}_{12}^t\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\right] = E_I^f\left[\mathbf{z}_2^t\left(\mathbf{I}_n - \mathbf{z}_1\left(\mathbf{z}_1^t\mathbf{z}_1\right)^{-1}\mathbf{z}_1^t\right)\mathbf{z}_2\right] \tag{2.6}$$
$$= E_I^f\left[\mathbf{v}_2^t\mathbf{v}_2\right] = (n-p+1)E_I^F[1],$$

$$E_I^f\left[\mathbf{T}_{11}^{-1}\mathbf{S}_{12}\sqrt{S_{22} - \mathbf{S}_{12}^t\mathbf{S}_{11}^{-1}\mathbf{S}_{12}}\right] = E_I^f\left[\mathbf{v}_1\sqrt{\mathbf{v}_2^t\mathbf{v}_2}\right] = \mathbf{0}. \tag{2.7}$$

On the basis of (2.5), (2.6) and (2.7), (2.4) is verified by induction. For $p = 2$, noting that $E_I^f[\mathbf{T}_{11}^t\mathbf{T}_{11}] = E_I^f[\mathbf{S}_{11}] = nE_I^F[1]$, we can easily see that $E_I^f[\mathbf{T}^t\mathbf{T}] = \operatorname{diag}(n+1, n-1)E_I^F[1]$. For $p \geq 3$, suppose that $E_I^f[\mathbf{T}_{11}^t\mathbf{T}_{11}] = \operatorname{diag}(n + (p-1) + 1 - 2i,\ i = 1, \ldots, p-1)E_I^F[1]$. Then from (2.5), $E_I^f[\mathbf{T}_{11}^t\mathbf{T}_{11} + \mathbf{T}_{21}^t\mathbf{T}_{21}] = \operatorname{diag}(n + p + 1 - 2i,\ i = 1, \ldots, p-1)E_I^F[1]$. Hence from (2.6) and (2.7), we get (2.4) and the first part of the proof is complete. The second part easily follows from the convexity of the loss function. □

For the assertion of the robustness of the improvement of $\hat{\boldsymbol{\Sigma}}^{SDS}$, the following lemma is essential.

LEMMA 2. *Let* $\mathbf{S} = \mathbf{HLH}^t$, $\mathbf{L} = \mathrm{diag}(l_1, \dots, l_p)$, $l_1 \geq \cdots \geq l_p$, *and consider the estimator* $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\phi}) = \mathbf{H}\,\mathrm{diag}(\phi_1(\mathbf{L}), \dots, \phi_p(\mathbf{L}))\mathbf{H}^t$. *Then under suitable conditions corresponding to those of Lemma* 1,

$$E^f_{\theta,\Sigma}\left[\mathrm{tr}\,\hat{\boldsymbol{\Sigma}}(\boldsymbol{\phi})\boldsymbol{\Sigma}^{-1}\right] = E^F_{\theta,\Sigma}\left[2\sum_{i\neq j}\frac{\phi_i(\mathbf{L})}{l_i - l_j} + 2\sum_i \frac{\partial\phi_i(\mathbf{L})}{\partial l_i} + (n-p-1)\sum_i \frac{\phi_i(\mathbf{L})}{l_i}\right].$$

This lemma is immediatley derived from Lemma 1 and the equation

$$\mathrm{tr}\left[\mathbf{D}_S\hat{\boldsymbol{\Sigma}}(\boldsymbol{\phi})\right] = \sum_{i\neq j}\phi_i(\mathbf{L})/(l_i - l_j) + \sum_i \partial\phi_i(\mathbf{L})/\partial l_i$$

as evaluated by Dey and Srinivasan (1985).

THEOREM 1. $\hat{\boldsymbol{\Sigma}}^{SDS}$ *is better than* $\hat{\boldsymbol{\Sigma}}^{JS}$ *uniformly for every unknown* $f(\cdot)$.

PROOF. Using (2.4) and Lemma 2, we can write the risk difference of estimators $\hat{\boldsymbol{\Sigma}}^{JS}$ and $\hat{\boldsymbol{\Sigma}}^{SDS}$ as

$$\begin{aligned}\Delta_2 &= R\left(\hat{\boldsymbol{\Sigma}}^{JS}, (\boldsymbol{\theta}, \boldsymbol{\Sigma}), f\right) - R\left(\hat{\boldsymbol{\Sigma}}^{SDS}, (\boldsymbol{\theta}, \boldsymbol{\Sigma}), f\right)\\ &= E^f_\Sigma[\mathrm{tr}\,\mathbf{DT}^t\boldsymbol{\Sigma}^{-1}\mathbf{T}] - E^f_\Sigma\left[\mathrm{tr}\,\mathbf{H}\,\mathrm{diag}(d_1 l_1, \dots, d_p l_p)\mathbf{H}^t\boldsymbol{\Sigma}^{-1}\right]\\ &= E^F_\Sigma[p] - E^F_\Sigma\left[2\sum_{i>j}\frac{d_i l_i - d_j l_j}{l_i - l_j} + 2\sum_i d_i + (n-p-1)\sum_i d_i\right].\end{aligned}\tag{2.8}$$

Using the equation

$$\frac{d_i l_i - d_j l_j}{l_i - l_j} = \frac{d_i - d_j}{l_i - l_j} l_i + d_j,$$

we can rewrite Δ_2 as

$$\begin{aligned}\Delta_2 &= -E^F_\Sigma\left[2\sum_{i>j}\frac{d_i - d_j}{l_i - l_j} l_i + \sum_i (n+p+1-2i)d_i - p\right]\\ &= -E^F_\Sigma\left[2\sum_{i>j}\frac{d_i - d_j}{l_i - l_j} l_i\right],\end{aligned}$$

since $\sum_{i>j} d_j = \sum_{i=1}^p \sum_{j=1}^{i-1} d_j = \sum_{j=1}^p \sum_{i=j+1}^p d_j = \sum_{j=1}^p (p-j) d_j$. For $i > j$, $d_i > d_j$ and $l_i < l_j$, so that we get that $\Delta_2 \geq 0$, and the proof is complete. □

Two major dominance results in estimation of the covariance matrix have thus been established to be robust in our sense. Also it can be verified that nonorder-preserving estimators are improved on by the corresponding order-preserving estimators in the ECD model when the function $f(\cdot)$ is nonincreasing. This result follows from the fact that Lemma 1 of Sheena and Takemura (1992) holds for nonincreasing function $f(\cdot)$. This demonstrates the inadmissibility of $\hat{\boldsymbol{\Sigma}}^{SDS}$ for $p \geq 2$ and every nonincreasing function $f(\cdot)$. □

In the ECD model, $n^{-1}\mathbf{S}$ is an unbiaesd estimator of $\boldsymbol{\Sigma}^* = E_{\Sigma}^{f}[n^{-1}\mathbf{S}] = E_I^F[1]\boldsymbol{\Sigma}$. By verifying each step of the above proofs, it can be shown that the robust dominance results obtained in this section still hold in the situation of estimation of $\boldsymbol{\Sigma}^*$.

APPENDIX

PROOF OF LEMMA 1. Before Haff (1979) established his identity for the Wishart distribution, Stein (1977a) had derived this identity by using the Stein identity which is technically very different from Haff's derivation. Using Stein's method, however, enables us to extend the so-called Haff's identity to the ECD model. We shall, therefore, more appropriately call it the Stein–Haff identity. A detailed proof of this identity using Stein's method for the normal model is given in Takemura (1991). The proof of the Lemma is now given in the following three steps, where without any loss of generality, we shall assume that $\boldsymbol{\theta} = \mathbf{0}$.

Let $h(\mathbf{S})$ be a scalar valued function of $\mathbf{S}$ and let $\boldsymbol{\Sigma} = \mathbf{I}_p$ for the $\boldsymbol{\Sigma}$ operated on $\mathbf{S}$ only. Noting that $\mathbf{S} = \mathbf{z}^t\mathbf{z}$ with $\mathbf{z} = (z_{ij})$, the same arguments as used in Cellier, Fourdrinier and Robert (1989) give that

$$
\begin{aligned}
&E_I^f\left[z_{ki}z_{kj}h(\mathbf{S})\right]\\
&\quad = \int\int z_{ki}z_{kj}h(\mathbf{S})\int|\boldsymbol{\Sigma}|^{-m/2}f\bigg(\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{x}^t\mathbf{x} + z_{kj}^2\\
&\qquad\qquad + \sum_{(a,b)\neq(k,j)} z_{ab}^2\bigg)\,d\mathbf{x}\,dz_{kj}\prod_{(a,b)\neq(k,j)} dz_{ab}\\
&\quad = \int\int \frac{\partial}{\partial z_{kj}}\{z_{ki}h(\mathbf{S})\}\int|\boldsymbol{\Sigma}|^{-m/2}F\bigg(\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{x}^t\mathbf{x} + z_{kj}^2\\
&\qquad\qquad + \sum_{(a,b)\neq(k,j)} z_{ab}^2\bigg)\,d\mathbf{x}\,dz_{kj}\prod_{(a,b)\neq(k,j)} dz_{ab},
\end{aligned}
\tag{A.1}
$$

which implies that

$$
\begin{aligned}
E_I^f\left[s_{ij}h(\mathbf{S})\right] &= \sum_{k=1}^n E_I^f\left[z_{ki}z_{kj}h(\mathbf{S})\right]\\
&= \sum_{k=1}^n E_I^F\left[\frac{\partial}{\partial z_{kj}}\{z_{ki}h(\mathbf{S})\}\right]\\
&= \sum_{k=1}^n E_I^F\left[\delta_{ij}h(\mathbf{S}) + z_{ki}\frac{\partial}{\partial z_{kj}}h(\mathbf{S})\right].
\end{aligned}
$$

Since $\partial s_{ab}/\partial z_{kj} = \delta_{ja} z_{kb} + \delta_{jb} z_{ka}$, we see that

$$\frac{\partial}{\partial z_{kj}} h(\mathbf{S}) = \sum_{a \geq b} \frac{\partial s_{ab}}{\partial z_{kj}} \frac{\partial h(\mathbf{S})}{\partial s_{ab}}$$

(A.2)
$$= \left(\sum_{j \geq b} z_{kb} \frac{\partial}{\partial s_{jb}} + \sum_{a \geq j} z_{ka} \frac{\partial}{\partial s_{aj}} \right) h(\mathbf{S})$$

$$= 2 \sum_{a} z_{ka} d_{aj} h(\mathbf{S}),$$

so that in the matrix form, we get the identity

(A.3)
$$E_I^f[\mathbf{S}h(\mathbf{S})] = E_I^F\big[nh(\mathbf{S})\mathbf{I}_p + 2\mathbf{S}\{\mathbf{D}_S h(\mathbf{S})\}\big],$$

where $\{\mathbf{D}_S h(\mathbf{S})\}_{ij} = d_{ij} h(\mathbf{S})$.

Let $\boldsymbol{\Sigma}$ be a $p \times p$ positive definite matrix and $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^t$. Then,

(A.4)
$$E_\Sigma^f[\mathbf{S}h(\mathbf{S})] = \mathbf{A} E_I^f\big[\mathbf{S}h(\mathbf{A}\mathbf{S}\mathbf{A}^t)\big]\mathbf{A}^t.$$

It is here noted that

(A.5)
$$d_{ij} h(\mathbf{A}\mathbf{S}\mathbf{A}^t) = \sum_{a \geq b} \frac{1}{2}(1 + \delta_{ij}) \frac{\partial(\mathbf{A}\mathbf{S}\mathbf{A}^t)_{ab}}{\partial s_{ij}} \frac{\partial h(\mathbf{A}\mathbf{S}\mathbf{A}^t)}{\partial(\mathbf{A}\mathbf{S}\mathbf{A}^t)_{ab}}$$

$$= \sum_{a, b} A_{ai} A_{bj} \tilde{d}_{ab} h(\mathbf{A}\mathbf{S}\mathbf{A}^t),$$

where $\mathbf{A} = (A_{ij})$ and

$$\tilde{d}_{ab} = \frac{1}{2}(1 + \delta_{ab}) \frac{\partial}{\partial(\mathbf{A}\mathbf{S}\mathbf{A}^t)_{ab}}.$$

Then (A.5) is rewritten in the matrix form as

(A.6)
$$\mathbf{D}_S h(\mathbf{A}\mathbf{S}\mathbf{A}^t) = \mathbf{A}^t\big\{\tilde{D}_S h(\mathbf{A}\mathbf{S}\mathbf{A}^t)\big\}\mathbf{A}.$$

Combining (A.3), (A.4) and (A.6) gives

$$E_\Sigma^f\big[\mathbf{S}\boldsymbol{\Sigma}^{-1} h(\mathbf{S})\big] = E_I^F\big[nh(\mathbf{A}\mathbf{S}\mathbf{A}^t)\mathbf{I}_p\big] + 2\mathbf{A} E_I^F\big[\mathbf{S}\{\mathbf{D}_S h(\mathbf{A}\mathbf{S}\mathbf{A}^t)\}\big]\mathbf{A}^{-1}$$

(A.7)
$$= E_I^F\big[nh(\mathbf{A}\mathbf{S}\mathbf{A}^t)\mathbf{I}_p\big] + 2E_I^F\big[\mathbf{A}\mathbf{S}\mathbf{A}^t\{\tilde{D}_S h(\mathbf{A}\mathbf{S}\mathbf{A}^t)\}\big]$$

$$= E_\Sigma^F\big[nh(\mathbf{S})\mathbf{I}_p + 2\mathbf{S}\{\mathbf{D}_S h(\mathbf{S})\}\big].$$

Let $\mathbf{H}(\mathbf{S})$ be a $p \times p$ matrix with the (i, j) element $h_{ji}(\mathbf{S})$ where for the function $h(\mathbf{S}) = h_{ji}(\mathbf{S})$, (A.7) is written as

$$E_\Sigma^f\left[\sum_a s_{ia} \sigma^{aj} h_{ji}(\mathbf{S})\right] = E_\Sigma^F\left[n\delta_{ij} h_{ji}(\mathbf{S}) + 2\sum_a s_{ia} d_{aj}\big(h_{ji}(\mathbf{S})\big)\right].$$

Taking the summation on i and j in the above equation, we obtain

$$E_\Sigma^f\big[\operatorname{tr}\mathbf{H}(\mathbf{S})\mathbf{S}\boldsymbol{\Sigma}^{-1}\big] = E_\Sigma^F\big[n\operatorname{tr}\mathbf{H}(\mathbf{S}) + 2\operatorname{tr}\mathbf{S}[\mathbf{D}_S\mathbf{H}(\mathbf{S})]\big].$$

Putting $\mathbf{G}(\mathbf{S}) = \mathbf{H}(\mathbf{S})\mathbf{S}$ gives

$$E_{\Sigma}^{f}\left[\operatorname{tr}\mathbf{G}(\mathbf{S})\boldsymbol{\Sigma}^{-1}\right] = E_{\Sigma}^{F}\left[n\operatorname{tr}\mathbf{G}(\mathbf{S})\mathbf{S}^{-1} + 2\operatorname{tr}\mathbf{S}\left[\mathbf{D}_S\{\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}\}\right]\right]. \tag{A.8}$$

Finally we evaluate the second term on the r.h.s. of (A.8). Note that

$$\left[\mathbf{D}_S\{\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}\}\right]_{ij} = \left(\{\mathbf{D}_S\mathbf{G}(\mathbf{S})\}\mathbf{S}^{-1}\right)_{ij} + \sum_{a,b} g_{ab}(\mathbf{S})d_{ia}s^{bj}. \tag{A.9}$$

Since $d\mathbf{S}^{-1} = -\mathbf{S}^{-1}(d\mathbf{S})\mathbf{S}^{-1}$, $d_{ia}s^{bj} = -2^{-1}(s^{ba}s^{ij} + s^{ib}s^{aj})$, so that

$$\begin{aligned}\sum_{a,b} g_{ab}(\mathbf{S})d_{ia}s^{bj} &= -\tfrac{1}{2}\sum_{a,b} g_{ab}(\mathbf{S})s^{ba}s^{ij} - \tfrac{1}{2}\sum_{a,b} g_{ab}(\mathbf{S})s^{ib}s^{aj}\\ &= -\tfrac{1}{2}S^{ij}\operatorname{tr}\left(\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}\right) - \tfrac{1}{2}\left(\mathbf{S}^{-1}\mathbf{G}(\mathbf{S})^t\mathbf{S}^{-1}\right)_{ij}.\end{aligned} \tag{A.10}$$

Combining (A.9) and (A.10) gives

$$\begin{aligned}\operatorname{tr}\mathbf{S}\left[\mathbf{D}_S\{\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}\}\right] &= \sum_{i,j} s_{ji}\left[\mathbf{D}_S\{\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}\}\right]_{ij}\\ &= \operatorname{tr}\mathbf{S}\left[\mathbf{D}_S\mathbf{G}(\mathbf{S})\right]\mathbf{S}^{-1} - \frac{1}{2}(\operatorname{tr}\mathbf{S}\mathbf{S}^{-1})\operatorname{tr}\left(\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}\right)\\ &\quad - \frac{1}{2}\operatorname{tr}\left(\mathbf{S}\mathbf{S}^{-1}\mathbf{G}(\mathbf{S})^t\mathbf{S}^{-1}\right)\\ &= \operatorname{tr}\left[\mathbf{D}_S\mathbf{G}(\mathbf{S})\right] - \frac{p+1}{2}\operatorname{tr}\mathbf{G}(\mathbf{S})\mathbf{S}^{-1}.\end{aligned} \tag{A.11}$$

From (A.8) and (A.11), the eliptically contoured version of the Haff identity follows. □

Acknowledgments. We are grateful to the Editor and the referee for their valuable comments and suggestions.

REFERENCES

BILODEAU, M. and KARIYA, T. (1989). Minimax estimators in the normal MANOVA model. *J. Multivariate Anal.* **28** 260–270.

CELLIER, D., FOURDRINIER, D. AND ROBERT, C. (1989). Robust shrinkage estimators of the location parameter for elliptically symmetric distributions. *J. Multivariate Anal.* **29** 39–52.

DEY, D. K. AND SRINIVASAN, C. (1985). Estimation of covariance matrix under Stein's loss. *Ann. Statist.* **13** 1581–1591.

HAFF, L. R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.* **9** 531–544.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press, Berkeley.

KUBOKAWA, T. (1998). The Stein phenomenon in simultaneous estimation: A review. In *Applied Statistical Science* **3** (S. E. Ahmed, M. Ahsanullah and B. K. Sinha, eds.) 143–173. NOVA, New York.

KUBOKAWA, T. and SRIVASTAVA, M. S. (1997). Robust improvements in estimation of mean and covariance matrices in elliptically contoured distribution. Discussion Paper Ser. 97-F-23, Faculty of Economics, Univ. Tokyo.

ROBERT, C. P. (1994). *The Bayesian Choice: A Decision-Theoretic Motivation*. Springer, New York.

SHEENA, Y. and TAKEMURA, A. (1992). Inadmissibility of non-order-preserving orthogonally invariant estimators of the covariance matrix in the case of Stein's loss. *J. Multivariate Anal.* **41** 117–131.

SRIVASTAVA, M. S. and BILODEAU, M. (1989). Stein estimation under elliptical distributions. *J. Multivariate Anal.* **28** 247–259.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press, Berkeley.

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

STEIN, C. (1977a). Estimating the covariance matrix. Unpublished manuscript.

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

SUGIURA, N. and ISHIBAYASHI, H. (1997). Reference prior Bayes estimator for bivariate normal covariance matrix with risk comparison. *Comm. Statist. Theory Methods* **26** 2203–2221.

TAKEMURA, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukuba J. Math.* **8** 367–376.

TAKEMURA, A. (1991). *Foundation of the Multivariate Statistical Inference*. Kyoritsu Press, Tokyo. (In Japanese.)

YANG, R. and BERGER, J. O. (1994). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* **22** 1195–1211.

UNIVERSITY OF TOKYO
FACULTY OF ECONOMICS
7-3-1 BUNKYO-KU
TOKYO 113
JAPAN

UNIVERSITY OF TORONTO
DEPARTMENT OF STATISTICS
100 ST. GEORGE ST.
TORONTO, ONTARIO M5S 3G3
CANADA