97-F-23

Robust Improvements in Estimation of Mean and Covariance Matrices in Elliptically Contoured Distribution

Tatsuya Kubokawa
University of Tokyo

M. S. Srivastava
University of Toronto

July 1997

Robust Improvements in Estimation of Mean and Covariance Matrices in Elliptically Contoured Distribution

T. KUBOKAWA and M.S. SRIVASTAVA
University of Tokyo and University of Toronto

This paper derives extended versions of 'Stein' and 'Haff' or more appropriately 'Stein-Haff' identities for elliptically contoured distribution (ECD) models. These identities are then used to establish the robustness of shrinkage estimators for the regression parameters in the multivariate linear regression model when the error matrix follows ECD model. Both invariant and non-invariant loss functions are considered in the above model as well as in the growth curve model. Also, the robustness of minimax estimators for the scale matrix in ECD models is established. Other results include the robustness, in a restricted ECD model, of the dominance results for the estimation of a scale with unknown locations and for the estimation of variance components in two components one-way mixed linear model with replicates.

AMS 1991 subject classifications. Primary 62C15; secondary 62F11, 62H12, 62J05.
Key words and phrases: Elliptically contoured distribution, robustness of improvement, multivariate linear model, growth curve model, regression coefficient matrix, covariance, shrinkage estimation, statistical decision theory, point estimation.

1 Introduction

Let us consider the multivariate linear regression model

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

where $\boldsymbol{y}$ is an $N \times p$ observed variable, $\boldsymbol{A}$ is an $N \times m$ known matrix with a full rank, $\boldsymbol{\beta}$ is an $m \times p$ unknown regression coefficient matrix, and $\boldsymbol{e}$ is an $N \times p$ error matrix. Assume that the error $\boldsymbol{e}$ has an elliptical density

$$|\Sigma|^{-N/2} f\left(\operatorname{tr} \Sigma^{-1} \boldsymbol{e}^t \boldsymbol{e}\right) \tag{1.2}$$

where Σ is a $p \times p$ unknown positive-definite matrix, $f(\cdot)$ is a nonnegative unknown function on the nonnegative real line, and $\boldsymbol{e}^t$ denotes the transpose of the matrix $\boldsymbol{e}$. The model (1.1) with (1.2) is called *Elliptically contoured model.* It may be noted that the density $f(\cdot)$ depends on N but for simplicity of notation this dependence is not shown.

In this paper, we consider the problem of estimating the coefficient matrix $\boldsymbol{\beta}$ and the scale matrix Σ for the elliptically contoured model (1.2) in a decision-theoretic set up. The performance of every estimator is evaluated in terms of a matrix or a scalar risk function. Beginning with the seminal work of Stein(1956) and James and Stein(1961), these problems have been extensively investigated in the statistical literature for the normal model; see Robert(1994) and Kubokawa(1997) for the vast literature in the normal model. The same, however, cannot be said for the elliptically contoured distribution model (1.2), hereafter referred to as ECD model. For example, no result exists in the literature in connection with the estimation of the scalar matrix Σ. Thus, it is not known whether the minimax estimator given by James and Stein(1961) and Dey and Srinivasan(1985) remain robust under the ECD model. We show in this paper that they remain robust.

Similarly, only partial results are available in estimating the coefficient matrix $\boldsymbol{\beta}$. For example, Srivastava and Bilodeau(1989) established the robustness of Stein estimator when the error matrix has the distribution of a scale mixture (with signed measure) of multivariate normal distribution and Cellier, Fourdrinier and Robert(1989) assumed that $p = 1$ and thus considered only the spherically symmetric model (SSD); for a survey and recent results on SSD models, see Brandwein and Strawderman(1990) and Cellier and Fourdrinier(1995a,b). In this paper, we provide a complete analogue of the results obtained by Bilodeau and Kariya(1989) and Konno(1990, 1991), for the ECD model. We also extend the results of Gleser(1987) and Honda(1991) for non-invariant loss functions. The double shrinkage estimator in the growth curve model given by Kariya, Konno and Strawderman(1996) is also shown to be robust for the ECD model.

Most results in the normal model employ the integration by parts approach of Stein(1973, 1981), known as 'Stein identity' and a related identity for Wishart distribution derived by Stein(1977a) and Haff(1979), known in the literature as the 'Haff identity'. We extend these identities to ECD models. Since our approach and proofs are based on Stein's method, we shall more appropriately call it Stein-Haff identity for the ECD model.

There are, however, some results which do not use the above two identities. This, for example, arises in estimating the scale parameter with unknown location parameters. Stein(1964) used the infinite series expansion of noncentral chi-square to prove the dominance result of his truncated estimator over the usual unbiased estimator in the normal model. For the ECD model we give a sufficient condition which, however, depends on f. Thus, we impose some conditions and obtain dominance result in a restricted class which include multivariate t-distribution.

Finally, we treat a one-way mixed linear model with r replicates and two variance components: 'within component' σ^2 and 'between component' σ_a^2. For $\sigma^{*2} = \sigma^2 + r\sigma_a^2$, $\sigma^2 \leq \sigma^{*2}$, the estimation of σ^2 and σ^{*2} are considered. In the normal model, unbiased estimators of σ^2 and σ^{*2} are improved upon by truncated procedures. In ECD model, it has, however, a different story. For estimation of σ^2, the robust improvement is established within a restricted class of ECD models, while the dominance result for the estimation of σ^{*2} is not robust. But a modification of a coefficient in the truncated procedure presents the robust improvement in the estimation of σ^{*2}.

The organization of the paper is as follows. In Section 2, we present Stein and Stein-Haff identities for the ECD model in two lemmas. Their proofs are also given. In Section 3, we consider the estimation of the coefficient matrix $\boldsymbol{\beta}$. Special cases of this problem, such as when $m = 1$ and the growth curve model, are given in subsections 3.1 and 3.4 respectively. Non-invariant loss functions are considered in subsection 3.3. Robust improvement in the estimation of the matrix Σ is considered in Section 4. Section 5 gives dominance results for the restricted ECD model in estimating the scale parameter with unknown location parameters. This section also includes the dominance results in a restricted ECD model for the estimation of variance components in the two components one-way mixed linear model with replicates.

2 An Extension of the Stein and Haff Identities to an Elliptically Contoured Distribution

We shall derive extended versions of Stein and Stein-Haff identities for the elliptically contoured distribution (ECD) model. These identities are heavily used in establishing the robust improvements of shrinkage estimators of the coefficient matrix and the scale matrix.

We begin with providing a canonical form of (1.1). let $\boldsymbol{H}$ be an $N \times N$ orthogonal matrix such that

$$\boldsymbol{HA} = \begin{pmatrix} (\boldsymbol{A}^t\boldsymbol{A})^{1/2} \\ \boldsymbol{0} \end{pmatrix}$$

and let $\boldsymbol{\theta} = (\boldsymbol{A}^t\boldsymbol{A})^{1/2}\boldsymbol{\beta}$. Let $\boldsymbol{x}$ and $\boldsymbol{z}$ be, respectively, $m \times p$ and $n \times p$ matrices such that

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} = H\boldsymbol{y} \quad \text{and} \quad n = N - m,$$

then the joint density of $\boldsymbol{x}$ and $\boldsymbol{z}$ has the form

$$|\Sigma|^{-N/2} f\left(\operatorname{tr}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta})^t(\boldsymbol{x}-\boldsymbol{\theta}) + \operatorname{tr}\Sigma^{-1}\boldsymbol{z}^t\boldsymbol{z}\right). \tag{2.1}$$

Denote $\boldsymbol{S} = \boldsymbol{z}^t\boldsymbol{z}$ and we treat the estimation issues of $\boldsymbol{\theta}$, Σ and their functions based on $\boldsymbol{x}$ and $\boldsymbol{S}$.

Let

$$F(x) = \frac{1}{2}\int_x^{+\infty} f(t)dt$$

and define

$$\begin{aligned} E^f_{\theta,\Sigma} h(\boldsymbol{x},\boldsymbol{z}) &= \int\int h(\boldsymbol{x},\boldsymbol{z})|\Sigma|^{-N/2} f\left(\operatorname{tr}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta})^t(\boldsymbol{x}-\boldsymbol{\theta}) + \operatorname{tr}\Sigma^{-1}\boldsymbol{z}^t\boldsymbol{z}\right) d\boldsymbol{x}d\boldsymbol{z}, \\ E^F_{\theta,\Sigma} h(\boldsymbol{x},\boldsymbol{z}) &= \int\int h(\boldsymbol{x},\boldsymbol{z})|\Sigma|^{-N/2} F\left(\operatorname{tr}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta})^t(\boldsymbol{x}-\boldsymbol{\theta}) + \operatorname{tr}\Sigma^{-1}\boldsymbol{z}^t\boldsymbol{z}\right) d\boldsymbol{x}d\boldsymbol{z}, \end{aligned}$$

where $h(\boldsymbol{x},\boldsymbol{z})$ is an integrable function. When there is no confusion we shall drop $\boldsymbol{\theta}$, Σ from the subscript in the above definitions. Let $\boldsymbol{x} = (x_{ij}) = (\boldsymbol{x}_1^t, \ldots, \boldsymbol{x}_m^t)^t$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^t, \ldots, \boldsymbol{\theta}_m^t)^t$. Then we get an identity corresponding to the Stein identity in the normal distribution.

Lemma 2.1. *Let $\boldsymbol{h} = (h_1, \ldots, h_p) : \boldsymbol{R}^p \to \boldsymbol{R}^p$ be a differentiable function and assume that*

(a) the absolute value of each element of $\{\boldsymbol{h}(\boldsymbol{x}_k)\}^t(\boldsymbol{x}_k - \boldsymbol{\theta})$ has a finite expectation,

(b) $\lim_{x_{kj}\to\pm\infty} h_i(x_{k1}, \ldots, x_{kj}, \ldots, x_{kp})F(x_{kj}^2 + a^2) = 0$ for $i = 1, \ldots, p$ and $j = 1, \ldots, p$.

Let $(\partial\{\boldsymbol{h}(\boldsymbol{x}_k)\}^t/\partial\boldsymbol{x}_k^t)_{ij} = \partial h_i(\boldsymbol{x}_k)/\partial x_{kj}$ and $(\boldsymbol{\nabla}_k\boldsymbol{h})_{ij} = \partial h_j/\partial x_{ki}$, where $(\boldsymbol{C})_{ij}$ designates the (i,j) element of matrix $\boldsymbol{C}$. Then

$$E^f_{\theta,\Sigma}[\{\boldsymbol{h}(\boldsymbol{x}_k)\}^t(\boldsymbol{x}_k - \boldsymbol{\theta}_k)] = E^F_{\theta,\Sigma}[\partial\{\boldsymbol{h}(\boldsymbol{x}_k)\}^t/\partial\boldsymbol{x}_k^t]\Sigma = E^F_{\theta,\Sigma}[(\boldsymbol{\nabla}_k\boldsymbol{h})^t\Sigma].$$

Proof. Letting $\boldsymbol{y}_k = (y_{k1}, \ldots, y_{kp}) = \boldsymbol{x}_k(\boldsymbol{C}^t)^{-1}$, $\boldsymbol{\xi}_k = (\xi_{k1}, \ldots, \xi_{kp}) = \boldsymbol{\theta}_k(\boldsymbol{C}^t)^{-1}$ and $\boldsymbol{C} = (c_{ij})$ for $\Sigma = \boldsymbol{CC}^t$, we observe that

$$E^f_{\theta,\Sigma}[\{\boldsymbol{h}(\boldsymbol{x}_k)\}^t(\boldsymbol{x}_k - \boldsymbol{\theta}_k)] = E^f_{\xi,I}[\{h(\boldsymbol{y}_k\boldsymbol{C}^t)\}^t(\boldsymbol{y}_k - \boldsymbol{\xi}_k)\boldsymbol{C}^t],$$

and from Cellier *et al.* (1989),

$$\begin{aligned} E^f_{\xi,I}[h_i(\boldsymbol{y}_k\boldsymbol{C}^t)(y_{kj} - \xi_{kj})] &= \int\cdot\int h_i(\boldsymbol{y}_k\boldsymbol{C}^t)(y_{kj} - \xi_{kj})f((y_{kj} - \xi_{kj})^2 + D)dy_{kj} \prod_{i\neq k, j\neq \ell} dy_{i\ell}d\boldsymbol{z} \\ &= \int\cdot\int \frac{\partial h_i(\boldsymbol{y}_k\boldsymbol{C}^t)}{\partial y_{kj}} F((y_{kj} - \xi_{kj})^2 + D)dy_{kj} \prod_{i\neq k, j\neq \ell} dy_{i\ell}d\boldsymbol{z}, \end{aligned} \tag{2.2}$$

where

$$D=\sum_{i=1,i\neq j}^{p}(y_{ki}-\xi_{ki})^2+\sum_{\ell=1,\ell\neq k}^{m}(\boldsymbol{y}_\ell-\boldsymbol{\xi}_\ell)(\boldsymbol{y}_\ell-\boldsymbol{\xi}_\ell)^t+\mathrm{tr}\,\boldsymbol{z}^t\boldsymbol{z}.$$

Hence,

$$\begin{aligned}E^f_{\xi,I}[h_i(\boldsymbol{y}_k\boldsymbol{C}^t)(y_{kj}-\xi_{kj})] &= E^F_{\xi,I}\left[\frac{\partial}{\partial y_{kj}}h_i(\boldsymbol{y}_k\boldsymbol{C}^t)\right]\\ &= E^F_{\xi,I}\left[\sum_\ell\frac{\partial h_i(\boldsymbol{y}_k\boldsymbol{C}^t)}{\partial(\boldsymbol{y}_k\boldsymbol{C}^t)_\ell}\times\frac{\partial(\boldsymbol{y}_k\boldsymbol{C}^t)_\ell}{\partial y_{kj}}\right]\\ &= E^F_{\theta,\Sigma}\left[\sum_\ell\frac{\partial h_i(\boldsymbol{y}_k)}{\partial y_{k\ell}}\times c_{\ell j}\right],\end{aligned}$$

giving the identity of Lemma 2.1. △

In the estimation of the mean vector and the covariance matrix, the Haff identity in the Wishart distribution is known to be very useful. This identity can be also extended to the ECD model. Let $\boldsymbol{G}(\boldsymbol{S})$ be a $p\times p$ matrix such that the (i,j) element $g_{ij}(\boldsymbol{S})$ is a function of $\boldsymbol{S}=(s_{ij})$ and denote

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

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 $\boldsymbol{S}=\sum_{i=1}^n\boldsymbol{z}_i^t\boldsymbol{z}_i$ for $n=N-m$, $\boldsymbol{z}=(\boldsymbol{z}_1^t,\ldots,\boldsymbol{z}_n^t)^t$ and $\boldsymbol{z}_k=(z_{k1},\ldots,z_{kp})$.

Lemma 2.2. *For $k=1,\ldots,n$ and $j=1,\ldots,p$, assume that $\boldsymbol{G}(\sum_{i=1}^n\boldsymbol{z}_i^t\boldsymbol{z}_i)$ is differentiable with respect to z_{kj} and that*

(a) $E^f_{\theta,\Sigma}[|\mathrm{tr}\,\{\boldsymbol{G}(\boldsymbol{S})\Sigma^{-1}|]$ is finite,

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

$$E^f_{\theta,\Sigma}\left[\mathrm{tr}\,\{\boldsymbol{G}(\boldsymbol{S})\Sigma^{-1}\}\right]=E^F_{\theta,\Sigma}\left[(n-p-1)\mathrm{tr}\,\left\{\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}\right\}+2\mathrm{tr}\,\{\boldsymbol{D}_S\boldsymbol{G}(\boldsymbol{S})\}\right].$$

Proof. Before Haff(1979) established his identity for the Wishart distribution, Stein(1977a) had derived this identity by using 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 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 $\boldsymbol{\theta}=\boldsymbol{0}$ and thus write $E^f_\Sigma[\cdot]$ for $E^f_{0,\Sigma}[\cdot]$.

1st step. Let $h(\boldsymbol{S})$ be a scalar valued function of $\boldsymbol{S}$ and let $\Sigma=\boldsymbol{I}_p$. Noting that $\boldsymbol{S}=\boldsymbol{z}^t\boldsymbol{z}$ with $\boldsymbol{z}=(z_{ij})$, the same arguments as in (2.2) gives that

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

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

$$\begin{aligned}
\frac{\partial}{\partial z_{kj}} h(\boldsymbol{S}) &= \sum_{a\geq b} \frac{\partial s_{ab}}{\partial z_{kj}} \frac{\partial h(\boldsymbol{S})}{\partial s_{ab}} \\
&= \sum_{a\geq b} (\delta_{ja} z_{kb} + \delta_{jb} z_{ka}) \frac{\partial h(\boldsymbol{S})}{\partial s_{ab}} \\
&= \left(\sum_{j\geq b} z_{kb} \frac{\partial}{\partial s_{jb}} + \sum_{a\geq j} z_{ka} \frac{\partial}{\partial s_{aj}} \right) h(\boldsymbol{S}) \\
&= 2 \sum_a \frac{1}{2}(1+\delta_{aj}) z_{ka} \frac{\partial}{\partial s_{aj}} h(\boldsymbol{S}) \\
&= 2 \sum_a z_{ka} d_{aj} h(\boldsymbol{S}),
\end{aligned} \tag{2.3}$$

so that

$$E_I^f [s_{ij} h(\boldsymbol{S})] = E_I^F \left[n\delta_{ij} h(\boldsymbol{S}) + 2 \sum_a \sum_k z_{ki} z_{ka} d_{aj} h(\boldsymbol{S}) \right],$$

which yields the matrix form

$$E_I^f [\boldsymbol{S} h(\boldsymbol{S})] = E_I^F \left[n h(\boldsymbol{S}) \boldsymbol{I}_p + 2\boldsymbol{S} \{ \boldsymbol{D} h(\boldsymbol{S}) \} \right], \tag{2.4}$$

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

<u>**2nd step.**</u> Let Σ be a $p \times p$ positive definite matrix and $\Sigma = \boldsymbol{A}\boldsymbol{A}^t$. Then,

$$\begin{aligned}
E_\Sigma^f [\boldsymbol{S} h(\boldsymbol{S})] &= E_I^f \left[\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) \right] \\
&= \boldsymbol{A} E_I^f \left[\boldsymbol{S} h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) \right] \boldsymbol{A}^t.
\end{aligned} \tag{2.5}$$

It is here noted that

$$\begin{aligned}
d_{ij} h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) &= \frac{1}{2}(1+\delta_{ij}) \frac{\partial}{\partial s_{ij}} h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) \\
&= \sum_{a\geq b} \frac{1}{2}(1+\delta_{ij}) \frac{\partial (\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t)_{ab}}{\partial s_{ij}} \cdot \frac{\partial h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t)}{\partial (\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t)_{ab}} \\
&= \sum_{a\geq b} A_{ai} A_{bj} \frac{\partial h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t)}{\partial (\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t)_{ab}} \\
&= \sum_{a,b} A_{ai} A_{bj} \tilde{d}_{ab} h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t),
\end{aligned} \tag{2.6}$$

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

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

(2.6) is rewritten in the matrix form as

$$\boldsymbol{D}_S h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) = \boldsymbol{A}^t \left\{ \tilde{D}_S h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) \right\} \boldsymbol{A}. \tag{2.7}$$

Combining (2.4), (2.5) and (2.7) gives

$$\begin{aligned}
E_\Sigma^f [\boldsymbol{S} h(\boldsymbol{S})] &= \boldsymbol{A} E_I^F \left[n h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) \boldsymbol{I}_p \right] \boldsymbol{A}^t + 2\boldsymbol{A} E_I^F \left[\boldsymbol{S} \{ \boldsymbol{D}_S h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) \} \right] \boldsymbol{A}^t \\
&= E_I^F \left[n\Sigma h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) \right] + 2 E_I^F \left[\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t \{ \tilde{D}_S h(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^t) \} \boldsymbol{A}\boldsymbol{A}^t \right] \\
&= E_\Sigma^F [n\Sigma h(\boldsymbol{S}) + 2\boldsymbol{S} \{ \boldsymbol{D}_S h(\boldsymbol{S}) \} \Sigma].
\end{aligned} \tag{2.8}$$

By multiplying Σ in (2.8) from the right, we get

$$E_{\Sigma}^{f}\left[\boldsymbol{S}\Sigma^{-1}h(\boldsymbol{S})\right] = E_{\Sigma}^{F}\left[nh(\boldsymbol{S})\boldsymbol{I}_p + 2\boldsymbol{S}\{\boldsymbol{D}_S h(\boldsymbol{S})\}\right]. \tag{2.9}$$

3rd step. Let $\boldsymbol{H}(\boldsymbol{S})$ be a $p \times p$ matrix with the (i,j) element $h_{ji}(\boldsymbol{S})$ where for the function $h(\boldsymbol{S}) = h_{ji}(\boldsymbol{S})$, (2.9) is written as

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

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

$$E_{\Sigma}^{f}\left[\mathrm{tr}\,\boldsymbol{H}(\boldsymbol{S})\boldsymbol{S}\Sigma^{-1}\right] E_{\Sigma}^{F}\left[n\mathrm{tr}\,\boldsymbol{H}(\boldsymbol{S}) + 2\mathrm{tr}\,\boldsymbol{S}[\boldsymbol{D}_S\boldsymbol{H}(\boldsymbol{S})]\right].$$

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

$$E_{\Sigma}^{f}[\mathrm{tr}\,\boldsymbol{G}(\boldsymbol{S})\Sigma^{-1}] = E_{\Sigma}^{F}[n\mathrm{tr}\,\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1} + 2\mathrm{tr}\,\boldsymbol{S}[\boldsymbol{D}_S\{\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}\}]]. \tag{2.10}$$

Finally we evaluate the second term on the r.h.s. of (2.10). Note that

$$\left[\boldsymbol{D}_S\{\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}\}\right]_{ij} = \left(\{\boldsymbol{D}_S\boldsymbol{G}(\boldsymbol{S})\}\boldsymbol{S}^{-1}\right)_{ij} + \sum_{a,b} g_{ab}(\boldsymbol{S})d_{ia}s^{bj}. \tag{2.11}$$

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

$$\sum_{a,b} g_{ab}(\boldsymbol{S})d_{ia}s^{bj} = -\frac{1}{2}\sum_{a,b} g_{ab}(\boldsymbol{S})s^{ba}s^{ij} - \frac{1}{2}\sum_{a,b} g_{ab}(\boldsymbol{S})s^{ib}s^{aj} \tag{2.12}$$

$$= -\frac{1}{2}s^{ij}\mathrm{tr}\left(\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}\right) - \frac{1}{2}\left(\boldsymbol{S}^{-1}\boldsymbol{G}(\boldsymbol{S})^t\boldsymbol{S}^{-1}\right)_{ij}. \tag{2.13}$$

Combing (2.11) and (2.12) gives

$$\begin{aligned}
\mathrm{tr}\,\boldsymbol{S}[\boldsymbol{D}_S\{\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}\}] &= \sum_{i,j} s_{ji}[\boldsymbol{D}_S\{\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}\}]_{ij} \\
&= \mathrm{tr}\,\boldsymbol{S}[\boldsymbol{D}_S\boldsymbol{G}(\boldsymbol{S})]\boldsymbol{S}^{-1} - \frac{1}{2}\left(\mathrm{tr}\,\boldsymbol{S}\boldsymbol{S}^{-1}\right)\mathrm{tr}\left(\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}\right) \\
&\quad -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{S}\boldsymbol{S}^{-1}\boldsymbol{G}(\boldsymbol{S})^t\boldsymbol{S}^{-1}\right) \qquad (2.14) \\
&= \mathrm{tr}\,[\boldsymbol{D}_S\boldsymbol{G}(\boldsymbol{S})] - \frac{p+1}{2}\mathrm{tr}\,\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}.
\end{aligned}$$

From (2.10) and (2.14), the elliptically contoured version of the Haff identity follows. $\triangle$

3 Robust Dominance Results for the Matrix Mean

3.1 One-dimensional case

We first discuss the case of $m = 1$ instructively. In the canonical form given by (2.1), we want to estimate $1 \times p$ vector $\boldsymbol{\theta}$ based on $\boldsymbol{x}$ and $\boldsymbol{z}$ (or $\boldsymbol{S}$) relative to the invariant loss function $(\boldsymbol{\delta} - \boldsymbol{\theta})\Sigma^{-1}(\boldsymbol{\delta} - \boldsymbol{\theta})^t$. Consider a class of shrinkage estimators

$$\boldsymbol{\delta}_\phi = \left(1 - \frac{\phi(\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t)}{\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t}\right)\boldsymbol{x}, \tag{3.1}$$

for absolutely continuous function $\phi(\cdot)$.

Proposition 3.1. *For $m = 1$ and $p \geq 3$, assume that $\phi(\cdot)$ is a nondecreasing function to the interval $(0, 2(p-2)/(N-p+2)]$. Then $\boldsymbol{\delta}_\phi$ dominates $\boldsymbol{\delta}_0 = \boldsymbol{x}$ uniformly for every unknown function $f(\cdot)$.*

Proof. The risk difference of two estimators $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}_\phi$ is written by

$$\begin{aligned}
\Delta_0 &= E^f_{\theta,\Sigma}\left[(\boldsymbol{\delta}_0-\boldsymbol{\theta})\Sigma^{-1}(\boldsymbol{\delta}_0-\boldsymbol{\theta})^t\right] - E^f_{\theta,\Sigma}\left[(\boldsymbol{\delta}_\phi-\boldsymbol{\theta})\Sigma^{-1}(\boldsymbol{\delta}_\phi-\boldsymbol{\theta})^t\right] \\
&= E^f_{\theta,\Sigma}\left[\frac{2\phi}{\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t}\boldsymbol{x}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta})^t - \frac{\phi^2}{(\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t)^2}\boldsymbol{x}\Sigma^{-1}\boldsymbol{x}^t\right] \\
&= 2E^F_{\theta,\Sigma}\left[\frac{(p-2)\phi}{\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t} + 2\phi'\right] - E^f_{\theta,\Sigma}\left[\frac{\phi^2}{(\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t)^2}\boldsymbol{x}\Sigma^{-1}\boldsymbol{x}^t\right],
\end{aligned} \tag{3.2}$$

where Lemma 2.1 was used in the third equality in (3.2). For evaluating the second term of the r.h.s. of the third equality, we make a scale transformation and get

$$E^f_{\theta,\Sigma}\left[\frac{\phi^2(\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t)}{(\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t)^2}\boldsymbol{x}\Sigma^{-1}\boldsymbol{x}^t\right] = E^f_{\xi,I}\left[\frac{\phi^2(\boldsymbol{x}(\boldsymbol{z}^t\boldsymbol{z})^{-1}\boldsymbol{x}^t)}{(\boldsymbol{x}(\boldsymbol{z}^t\boldsymbol{z})^{-1}\boldsymbol{x}^t)^2}\boldsymbol{x}\boldsymbol{x}^t\right], \tag{3.3}$$

where we use the same notations for $\boldsymbol{x}$ and $\boldsymbol{z}$ (or $\boldsymbol{S}$) after the transformation. Let $\boldsymbol{Q}_1$ be a $p \times p$ orthogonal matrix such that $\boldsymbol{Q}_1\boldsymbol{x}^t = (\sqrt{\boldsymbol{x}\boldsymbol{x}^t}, 0, \ldots, 0)^t$, and let $(\boldsymbol{y}_1^t, \boldsymbol{y}_2^t)^t = \boldsymbol{Q}_1\boldsymbol{z}^t$ with $1 \times (N-1)$ vector $\boldsymbol{y}_1$ and $(p-1) \times (N-1)$ matrix $\boldsymbol{y}_2$. Then,

$$\boldsymbol{x}(\boldsymbol{z}^t\boldsymbol{z})^{-1}\boldsymbol{x}^t = \boldsymbol{x}\boldsymbol{x}^t/\{\boldsymbol{y}_1(\boldsymbol{I}_{N-1} - \boldsymbol{y}_2^t(\boldsymbol{y}_2\boldsymbol{y}_2^t)^{-1}\boldsymbol{y}_2)\boldsymbol{y}_1^t\}.$$

Further, let $\boldsymbol{Q}_2$ be an $(N-1) \times (N-1)$ orthogonal matrix such that $\boldsymbol{Q}_2\boldsymbol{y}_2^t = (\boldsymbol{0}, (\boldsymbol{y}_2\boldsymbol{y}_2^t)^{1/2})^t$, and let

$$(\boldsymbol{u}_1, \boldsymbol{u}_2) = \boldsymbol{y}_1\boldsymbol{Q}_2^t,$$

with $1 \times (N-p)$ vector $\boldsymbol{u}_1$ and $1 \times (p-1)$ vector $\boldsymbol{u}_2$. Then $\boldsymbol{x}(\boldsymbol{z}^t\boldsymbol{z})^{-1}\boldsymbol{x}^t = \boldsymbol{x}\boldsymbol{x}^t/\boldsymbol{u}_1\boldsymbol{u}_1^t$ and $\mathrm{tr}\,\boldsymbol{z}^t\boldsymbol{z} = \boldsymbol{u}_1\boldsymbol{u}_1^t + \boldsymbol{u}_2\boldsymbol{u}_2^t + \mathrm{tr}\,(\boldsymbol{y}_2\boldsymbol{y}_2^t)$, which gives that for $\boldsymbol{\xi} = \boldsymbol{\theta}\Sigma^{-1/2}$,

$$\begin{aligned}
&E^f_{\xi,I}\left[\frac{\phi^2(\boldsymbol{x}(\boldsymbol{z}^t\boldsymbol{z})^{-1}\boldsymbol{x}^t)}{(\boldsymbol{x}(\boldsymbol{z}^t\boldsymbol{z})^{-1}\boldsymbol{x}^t)^2}\boldsymbol{x}\boldsymbol{x}^t\right] \\
&= \int\!\!\int\!\!\int\!\!\int \frac{(\boldsymbol{u}_1\boldsymbol{u}_1^t)^2}{\boldsymbol{x}\boldsymbol{x}^t}\phi^2\left(\frac{\boldsymbol{x}\boldsymbol{x}^t}{\boldsymbol{u}_1\boldsymbol{u}_1^t}\right) \\
&\qquad \times f\left(\mathrm{tr}\,(\boldsymbol{x}-\boldsymbol{\xi})^t(\boldsymbol{x}-\boldsymbol{\xi}) + \boldsymbol{u}_1\boldsymbol{u}_1^t + \boldsymbol{u}_2\boldsymbol{u}_2^t + \mathrm{tr}\,(\boldsymbol{y}_2\boldsymbol{y}_2^t)\right) d\boldsymbol{x}\,d\boldsymbol{u}_1\,d\boldsymbol{u}_2\,d\boldsymbol{y}_2 \\
&= E^F_{\xi,I}\left[(N-p+2)\frac{\boldsymbol{u}_1\boldsymbol{u}_1^t}{\boldsymbol{x}\boldsymbol{x}^t}\phi^2 - 4\phi\phi'\right] \\
&= E^F_{\xi,\Sigma}\left[\frac{N-p+2}{\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t}\phi^2 - 4\phi\phi'\right],
\end{aligned} \tag{3.4}$$

where Lemma 2.1 was used again with respect to $\boldsymbol{u}_1$ in the second equality of (3.4). Combining (3.2) and (3.4) shows that

$$\Delta_0 = E^F_{\theta,\Sigma}\left[\frac{2(p-2) - (N-p+2)\phi}{\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t} \times \phi + 4\phi'(1+\phi)\right],$$

which is guaranteed to be nonnegative under the conditions of Proposition 3.1. $\triangle$

3.2 Multi-dimensional case

A major interest in this section is to show that the robustness of dominance results in the shrinkage estimation is insured in the more general setting (1.1). Consider estimators of the general form

$$\delta(\boldsymbol{G}) = \boldsymbol{x} + \boldsymbol{G}(\boldsymbol{x}, \boldsymbol{S}) = \boldsymbol{x} + \boldsymbol{G} \tag{3.5}$$

and let $\boldsymbol{x} = (x_{ij}) = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_p^*) = (\boldsymbol{x}_1^t, \ldots, \boldsymbol{x}_m^t)^t$ and $\boldsymbol{G} = (g_{ij}) = (\boldsymbol{G}_1^*, \ldots, \boldsymbol{G}_p^*) = (\boldsymbol{G}_1^t, \ldots, \boldsymbol{G}_m^t)^t$. Note that the (k, ℓ) element of $p \times p$ matrix $\partial \boldsymbol{G}_i / \partial \boldsymbol{x}_j$ is $\partial g_{i\ell} / \partial x_{jk}$, while the (k, ℓ) element of $m \times m$ matrix $\partial \boldsymbol{G}_i^* / \partial \boldsymbol{x}_j^*$ is $\partial g_{ik} / \partial x_{j\ell}$. Two types of criteria for comparing estimators are treated:

$$\begin{aligned} \boldsymbol{R}_1(\boldsymbol{\delta}, (\boldsymbol{\theta}, \Sigma), f) &= E^f_{\theta,\Sigma}[(\boldsymbol{\delta} - \boldsymbol{\theta})\Sigma^{-1}(\boldsymbol{\delta} - \boldsymbol{\theta})^t], \quad m \times m \\ R_2(\boldsymbol{\delta}, (\boldsymbol{\theta}, \Sigma), f) &= E^f_{\theta,\Sigma}[\mathrm{tr}\, (\boldsymbol{\delta} - \boldsymbol{\theta})\Sigma^{-1}(\boldsymbol{\delta} - \boldsymbol{\theta})^t], \end{aligned}$$

where $\boldsymbol{R}_1(\boldsymbol{\delta}, (\boldsymbol{\theta}, \Sigma), f)$ is an $m \times m$ matrix and we say that $\boldsymbol{\delta}_1$ is better than $\boldsymbol{\delta}_2$ in terms of $\boldsymbol{R}_1$ if $\boldsymbol{R}_1(\boldsymbol{\delta}_2, (\boldsymbol{\theta}, \Sigma), f) - \boldsymbol{R}_1(\boldsymbol{\delta}_1, (\boldsymbol{\theta}, \Sigma), f)$ is non-negative definite for every $(\boldsymbol{\theta}, \Sigma)$ and the positive definiteness holds for some $(\boldsymbol{\theta}, \Sigma)$.

In this general setup, Bilodeau and Kariya(1989) derived a condition for $\boldsymbol{\delta}(\boldsymbol{G})$ given by (3.5) to dominate $\boldsymbol{x}$ through an unbiased estimator of the risk matrix $\boldsymbol{R}_1(\boldsymbol{\delta}(\boldsymbol{G}), (\boldsymbol{\theta}, \Sigma), f)$ in the normal distribution. Their general result can be extended to the ECD model based on Lemmas 2.1 and 2.2.

Theorem 3.1. *Assume that $\boldsymbol{G}(\boldsymbol{x}, \boldsymbol{S})$ satisfies the conditions of Lemmas 2.1 and 2.2. Then the (i, j) element of the $\boldsymbol{R}_1$-risk difference of the estimators $\boldsymbol{x}$ and $\delta(\boldsymbol{G})$ given by (3.5) is evaluated as*

$$\begin{aligned} (\Delta_1)_{ij} &= (\boldsymbol{R}_1(\boldsymbol{\delta}(\boldsymbol{G}), (\boldsymbol{\theta}, \Sigma), f) - \boldsymbol{R}_1(\boldsymbol{x}, (\boldsymbol{\theta}, \Sigma), f))_{ij} \\ &= E^F_{\theta,\Sigma}\left[\mathrm{tr}\, \nabla_i \boldsymbol{G}_j + \mathrm{tr}\, \nabla_j \boldsymbol{G}_i + (n - p - 1)\boldsymbol{G}_i \boldsymbol{S}^{-1} \boldsymbol{G}_j^t + 2\mathrm{tr}\, \boldsymbol{D}_S \boldsymbol{G}_i^t \boldsymbol{G}_j\right]. \end{aligned}$$

Proof. Observe that

$$\Delta_1 = E^f\left[(\boldsymbol{x} - \boldsymbol{\theta})\Sigma^{-1}\boldsymbol{G}^t + \boldsymbol{G}\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\theta})^t + \boldsymbol{G}\Sigma^{-1}\boldsymbol{G}^t\right].$$

Using Lemma 2.2, the (i, j) element of $E^f[\boldsymbol{G}_i \Sigma^{-1} \boldsymbol{G}_j^t]$ is evaluated as

$$\begin{aligned} E^f\left[\boldsymbol{G}_i \Sigma^{-1} \boldsymbol{G}_j^t\right] &= E^f\left[\mathrm{tr}\, \boldsymbol{G}_j^t \boldsymbol{G}_i \Sigma^{-1}\right] \\ &= E^F\left[(n - p - 1)\boldsymbol{G}_i \boldsymbol{S}^{-1} \boldsymbol{G}_j^t + 2\mathrm{tr}\, \left[\boldsymbol{D}_S \boldsymbol{G}_j^t \boldsymbol{G}_i\right]\right], \end{aligned}$$

which yields

$$\left(E^f\left[\boldsymbol{G}\Sigma^{-1}\boldsymbol{G}^t\right]\right)_{ij} = E^F\left[(n - p - 1)\boldsymbol{G}_i \Sigma^{-1} \boldsymbol{G}_j^t + 2\mathrm{tr}\, \boldsymbol{D}_S \boldsymbol{G}_i^t \boldsymbol{G}_j\right].$$

Also from Lemma 2.1, the (i, j) element of $E[(\boldsymbol{x} - \boldsymbol{\theta})\Sigma^{-1}\boldsymbol{G}^t]$ is rewritten as

$$\begin{aligned} E^f\left[(\boldsymbol{x}_i - \boldsymbol{\theta}_i)\Sigma^{-1}\boldsymbol{G}_j^t\right] &= \mathrm{tr}\, \Sigma^{-1} E^f\left[\boldsymbol{G}_j^t(\boldsymbol{x}_i - \boldsymbol{\theta}_i)\right] \\ &= \mathrm{tr}\, \Sigma^{-1} E^F\left[(\nabla_i \boldsymbol{G}_j)^t\, \Sigma\right] \\ &= E^F\left[\mathrm{tr}\, \nabla_i \boldsymbol{G}_j\right], \end{aligned}$$

so that

$$\left(E^f\left[(\boldsymbol{x} - \boldsymbol{\theta})\Sigma^{-1}\boldsymbol{G}^t\right]\right)_{ij} = E^F\left[\mathrm{tr}\, \nabla_i \boldsymbol{G}_j\right],$$

and the proof of Theorem 3.1 is complete. △

The following result for the R_2-risk difference is a direct consequence of Theorem 3.1.

Corollary 3.1. *Under the same assumptions as in Theorem 3.1, the R_2-risk difference of the estimators $\boldsymbol{x}$ and $\boldsymbol{\delta}(\boldsymbol{G})$ is evaluated as*

$$\begin{aligned}\Delta_2 &= R_2(\boldsymbol{\delta}(\boldsymbol{G}),(\boldsymbol{\theta},\Sigma),f) - R_2(\boldsymbol{x},(\boldsymbol{\theta},\Sigma),f) = \mathrm{tr}\,\Delta_1 \\ &= E^F_{\theta,\Sigma}\left[2\sum_{i=1}^{m}\mathrm{tr}\,\{\nabla_i \boldsymbol{G}_i\} + (n-p-1)\mathrm{tr}\,\{\boldsymbol{G}^t\boldsymbol{G}\boldsymbol{S}^{-1}\} + 2\mathrm{tr}\,\{\boldsymbol{D}_S\boldsymbol{G}^t\boldsymbol{G}\}\right].\end{aligned}$$

Bilodeau and Kariya(1989) derived several classes of improved estimators under very general conditions. In particular, it includes Konno's(1990) invariant class of estimators. These estimators are invariant under the group of transformations $(\boldsymbol{x},\boldsymbol{S}) \to (\boldsymbol{O}\boldsymbol{x}\boldsymbol{C}, \boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})$ where $\boldsymbol{O}$ is an $m\times m$ orthogonal matrix and $\boldsymbol{C}$ is a $p \times p$ nonsingular matrix, and are obtained as follows. Let $\boldsymbol{F} = (f_1,\ldots,f_{\min(p,m)})^t$ be a vector of ordered eigen values $f_1 \geq \cdots \geq f_{\min(p,m)}$ of the matrix $\boldsymbol{x}^t\boldsymbol{x}\boldsymbol{S}^{-1}$ defined as

$$\begin{aligned}\boldsymbol{R}_1^t\boldsymbol{S}\boldsymbol{R}_1 = \boldsymbol{I}_p, \quad \mathrm{diag}\,(\boldsymbol{F}) = \boldsymbol{R}_1^t\boldsymbol{x}^t\boldsymbol{x}\boldsymbol{R}_1, &\qquad \text{if} \quad m > p, \\ \boldsymbol{R}_2^t\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t\boldsymbol{R}_2 = \mathrm{diag}\,(\boldsymbol{F}) &\qquad \text{if} \quad m \leq p,\end{aligned}$$

where $\mathrm{diag}\,(\boldsymbol{F})$ is a diagonal matrix with diagonal elements as the ordered eigen values $f_1 \geq \cdots \geq f_{\min(p,m)}$, and $\boldsymbol{R}_1$, $\boldsymbol{R}_2$ are, respectively, $p\times p$, $m\times m$ nonsingular matrices. Let $h(\boldsymbol{F})$ be an absolutely continuous positive scalar function of the vector $\boldsymbol{F}$. Then Konno's(1990) estimator for the scalar loss function is given by

$$\begin{aligned}\boldsymbol{\delta}^{KN}(h) &= \boldsymbol{x}\left(\boldsymbol{I}_p + \boldsymbol{R}_1\boldsymbol{H}_1(\boldsymbol{F})\boldsymbol{R}_1^{-1}\right) &\quad \text{if} \quad m > p, \\ &= \left(\boldsymbol{I}_m + \boldsymbol{R}_2\boldsymbol{H}_2(\boldsymbol{F})\boldsymbol{R}_2^{-1}\right)\boldsymbol{x} &\quad \text{if} \quad m \leq p,\end{aligned}$$

where $\boldsymbol{H}_1(\boldsymbol{F}) = \mathrm{diag}\,(h_1(\boldsymbol{F}),\ldots,h_p(\boldsymbol{F}))$, $h_i(\boldsymbol{F}) = \partial h(\boldsymbol{F})/\partial f_i$, $i = 1,\ldots,p$, and $\boldsymbol{H}_2(\boldsymbol{F}) = \mathrm{diag}\,(h_1(\boldsymbol{F}),\ldots,h_m(\boldsymbol{F}))$. For example, if we choose $h(\boldsymbol{F}) = c_1\log(\pi_k f_k) + c_2\log(\sum_k f_k)$, then we obtain Efron-Morris'(1976) type of estimators given by Konno(1990):

$$\begin{aligned}\boldsymbol{\delta}^{EM}(c_1,c_2) &= \boldsymbol{x}\left[\boldsymbol{I}_p + c_1(\boldsymbol{x}^t\boldsymbol{x})^{-1}\boldsymbol{S} + c_2\frac{1}{\mathrm{tr}\,(\boldsymbol{x}^t\boldsymbol{x})\boldsymbol{S}^{-1}}\boldsymbol{I}_p\right] &\quad \text{for} \quad m \geq p+2, \\ &= \left[\boldsymbol{I}_m + c_1(\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t)^{-1} + c_2\frac{1}{\mathrm{tr}\,\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t}\boldsymbol{I}_m\right]\boldsymbol{x} &\quad \text{for} \quad p \geq m+2,\end{aligned}$$

where c_1 and c_2 are given by $c_1 = -(m-p-1)/(n+p+1)$, $c_2 = -(p^2+p-2)/(n-p+3)$ for $m \geq p+2$, and $c_1 = -(p-m-1)/(n+2m-p+1)$, $c_2 = -(m^2+m-2)/(n-p+3)$ for $p \geq m+2$. Another possible estimator $\boldsymbol{\delta}^{KN*}$ is given by putting $\boldsymbol{H}_1(\boldsymbol{F}) = \mathrm{diag}\,(d_1/f_1,\ldots,d_p/f_p)$ for $m > p$ and $\boldsymbol{H}_2(\boldsymbol{F}) = \mathrm{diag}\,(d_1/f_1,\ldots,d_m/f_m)$ for $m \leq p$ where $d_k = (m+p-2k-1)/(n-p+2k+1)$. Our Corollary 3.1 implies that these estimators $\boldsymbol{\delta}^{EM}(c_1,c_2)$ and $\boldsymbol{\delta}^{KN*}$ have robust improvements over the crude Efron-Morris estimator $\boldsymbol{\delta}^{EM}(c_1,0)$, being better than the least squares estimator $\boldsymbol{x}$. Similarly, other classes of estimators with respect to both kinds of loss functions given by Bilodeau and Kariya(1989) are also robust from our Theorem 3.1 and Corollary 3.1.

3.3 Developments under a non-invariant loss

It is of another great interest to investigate whether the robust improvements of shrinkage procedures remain true still for noninvariant loss functions. Suppose that estimator $\boldsymbol{\delta}$ is evaluated in terms of the following risk functions relative to the noninvariant loss functions:

$$\begin{aligned} \boldsymbol{R}_{N1}(\boldsymbol{\delta}, (\boldsymbol{\theta}, \Sigma), f) &= E^f_{\theta,\Sigma}\left[(\boldsymbol{\delta}-\boldsymbol{\theta})(\boldsymbol{\delta}-\boldsymbol{\theta})^t\right] \; : \; m \times m \\ R_{N2}(\boldsymbol{\delta}, (\boldsymbol{\theta}, \Sigma), f) &= E^f_{\theta,\Sigma}\left[\mathrm{tr}\,(\boldsymbol{\delta}-\boldsymbol{\theta})(\boldsymbol{\delta}-\boldsymbol{\theta})^t\right]. \end{aligned}$$

Gleser(1986) successfully developed improved shrinkage estimators for $m = 1$ in the normal distribution, and Honda(1991) extended it to the multivariate regression model. When we want to address their robustness, the essential part of it is to evaluate the cross term between $\boldsymbol{x}$ and the shrinkage function, in which Gleser(1986) utilized both of the Stein and Haff identities. For the purpose, we use the notation

$$E^{Df}_{\theta,\Sigma}[h(\boldsymbol{x},\boldsymbol{z})] = \int\int h(\boldsymbol{x},\boldsymbol{z})|\Sigma|^{-N/2}\left\{-2f'\left(\mathrm{tr}\,\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta})^t(\boldsymbol{x}-\boldsymbol{\theta}) + \mathrm{tr}\,\Sigma^{-1}\boldsymbol{z}^t\boldsymbol{z}\right)\right\} d\boldsymbol{x}d\boldsymbol{z}. \quad (3.6)$$

Following Gleser(1986) and Honda(1991), let $\boldsymbol{H} = (\boldsymbol{h}^t_1, \dots, \boldsymbol{h}^t_m)^t$ be an $m \times p$ matrix function of $\boldsymbol{x}$ and $\boldsymbol{S}$ and define $\boldsymbol{R} = (r_{ij}) = (\boldsymbol{r}^t_1, \dots, \boldsymbol{r}^t_m)^t$ $(m \times p)$ from $\boldsymbol{H}$ by

$$r_{ij} = \frac{\partial}{\partial s_{jj}}(\boldsymbol{h}_i(\boldsymbol{x},\boldsymbol{S})\boldsymbol{S})_j + \frac{1}{2}\sum_{k\neq j}\frac{\partial}{\partial s_{kj}}(\boldsymbol{h}_i(\boldsymbol{x},\boldsymbol{S})\boldsymbol{S})_k, \quad i = 1,\dots,m, \; j = 1,\dots,p,$$

where for vector $\boldsymbol{a}$, $(\boldsymbol{a})_k$ designates the kth element of $\boldsymbol{a}$. Then the estimator we consider is of the form

$$\boldsymbol{\delta}^{GL}_H = \boldsymbol{x} - \boldsymbol{T} \quad (3.7)$$

where

$$\boldsymbol{T} = (\boldsymbol{t}^t_1, \dots, \boldsymbol{t}^t_m)^t = \boldsymbol{H} + \frac{2}{n-p-1}\boldsymbol{R}.$$

The $\boldsymbol{R}_{N1}$-risk difference of $\boldsymbol{x}$ and $\boldsymbol{\delta}^{GL}_H$ is written by

$$\begin{aligned} \Delta_{N1} &= \boldsymbol{R}_{N1}(\boldsymbol{x}, (\boldsymbol{\theta}, \Sigma), f) - \boldsymbol{R}_{N1}(\boldsymbol{x}, (\boldsymbol{\theta}, \Sigma), f) \\ &= E^f_{\theta,\Sigma}\left[(\boldsymbol{x}-\boldsymbol{\theta})\boldsymbol{T}^t + \boldsymbol{T}(\boldsymbol{x}-\boldsymbol{\theta})^t - \boldsymbol{T}\boldsymbol{T}^t\right]. \end{aligned}$$

Using the notation (3.6) and Lemma 2.2, we see that

$$\begin{aligned} &\left(E^{Df}_{\theta,\Sigma}\left[\boldsymbol{H}\boldsymbol{S}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta})^t\right]\right)_{ij} \\ &= E^{Df}_{\theta,\Sigma}\left[\mathrm{tr}\,\{(\boldsymbol{x}_j-\boldsymbol{\theta}_j)^t\boldsymbol{h}_i\boldsymbol{S}\Sigma^{-1}\}\right] \quad (3.8) \\ &= E^f_{\theta,\Sigma}\left[(n-p-1)\mathrm{tr}\,\{(\boldsymbol{x}_j-\boldsymbol{\theta}_j)^t\boldsymbol{h}_i\} + 2\mathrm{tr}\,\boldsymbol{D}_S\{(\boldsymbol{x}_j-\boldsymbol{\theta}_j)^t\boldsymbol{h}_i\boldsymbol{S}\}\right] \\ &= (n-p-1)E^f_{\theta,\Sigma}\left[\mathrm{tr}\,(\boldsymbol{x}_j-\boldsymbol{\theta}_j)^t\boldsymbol{t}_i\right]. \end{aligned}$$

From Lemma 2.1, on the other hand,

$$E^{Df}_{\theta,\Sigma}\left[\mathrm{tr}\,\{(\boldsymbol{x}_j-\boldsymbol{\theta}_j)^t\boldsymbol{h}_i(\boldsymbol{x},\boldsymbol{S})\boldsymbol{S}\Sigma^{-1}\}\right] = E^f_{\theta,\Sigma}\left[\mathrm{tr}\,\nabla_j\{\boldsymbol{h}_i(\boldsymbol{x},\boldsymbol{S})\boldsymbol{S}\}\right]. \quad (3.9)$$

Combining (3.8) and (3.9), we get the required representation of Δ_{N1}.

Theorem 3.2. *Let f and $\boldsymbol{H}$ be differentiable, and assume that $E^{Df}_{\theta,\Sigma}[|\mathrm{tr}\,\{(\boldsymbol{x}_j-\boldsymbol{\theta}_j)^t\boldsymbol{h}_i(\boldsymbol{x},\boldsymbol{S})\boldsymbol{S}\Sigma^{-1}|]$ $< \infty$ for every i, j, and that $\lim_{z_{kj}\to\pm\infty}\boldsymbol{h}_i(\boldsymbol{x},\boldsymbol{S})f(z^2_{kj}+a^2) = \boldsymbol{0}$ and $\lim_{x_{kj}\to\pm\infty}\boldsymbol{h}_i(\boldsymbol{x},\boldsymbol{S})f(x^2_{kj}+$*

$a^2) = \mathbf{0}$ *for every* i, j, k *where* x_{kj} *and* z_{jk} *are, respectively, defined above Lemmas 2.1 and 2.2. Then the* (i,j) *element of* Δ_{N1} *is given by*

$$(\Delta_{N1})_{ij} = E^f_{\theta,\Sigma}\left[\frac{1}{n-p-1}\{\mathrm{tr}\,\nabla_j(\boldsymbol{h}_i\boldsymbol{S})\mathrm{tr}\,\nabla_i(\boldsymbol{h}_j\boldsymbol{S})\} - \boldsymbol{t}_i\boldsymbol{t}_j^t\right]. \tag{3.10}$$

Solutions for the nonnegativeness of the risk difference (Δ_{ij}) are proposed by Gleser(1986) as

$$\boldsymbol{H} = \frac{c\, b(\boldsymbol{S})}{\mathrm{tr}\,\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t}\boldsymbol{x}\boldsymbol{S}^{-1} \quad \text{for} \quad p \ge 3 \quad \text{and} \quad 0 < c \le \frac{2(p-2)(n-p-1)}{(n-p+3)^2},$$

where two common choices of $b(\boldsymbol{S})$ are $b_1(\boldsymbol{S}) = (\mathrm{tr}\,\boldsymbol{S}^{-1})^{-1}$ and $b_2(\boldsymbol{S}) = \lambda_{min}(\boldsymbol{S})$ for the minimum eigenvalue $\lambda_{min}(\boldsymbol{S})$ of $\boldsymbol{S}$. These choices yield, respectively,

$$\begin{aligned}
\boldsymbol{R}_1 &= \frac{cb_1(\boldsymbol{S})}{\mathrm{tr}\,\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t}\left\{\frac{1}{\mathrm{tr}\,\boldsymbol{S}^{-1}}\boldsymbol{x}\boldsymbol{S}^{-2} + \frac{1}{\mathrm{tr}\,\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t}(\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t)\boldsymbol{x}\boldsymbol{S}^{-1}\right\}, \\
\boldsymbol{R}_2 &= \frac{cb_2(\boldsymbol{S})}{\mathrm{tr}\,\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t}\left\{\frac{1}{\lambda_{min}(\boldsymbol{S})}\boldsymbol{x}\boldsymbol{g}\boldsymbol{g}^t + \frac{1}{\mathrm{tr}\,\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t}(\boldsymbol{x}\boldsymbol{S}^{-1}\boldsymbol{x}^t)\boldsymbol{x}\boldsymbol{S}^{-1}\right\},
\end{aligned}$$

where $\boldsymbol{g}$ denotes the eigenvector of $\boldsymbol{S}$ corresponding to $\lambda_{min}(\boldsymbol{S})$ such that $\boldsymbol{g}^t\boldsymbol{g} = 1$. The robust improvement for the scalar risk R_{N2} follows from the above arguments and Honda(1991).

3.4 Extensions to a growth curve model

We consider extending the robust dominance results to, more generally, a growth curve (or GMANOVA) model

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\beta}\boldsymbol{B} + \boldsymbol{e}, \tag{3.11}$$

where $\boldsymbol{y}$ is an $N \times p$ observed variable, $\boldsymbol{A}$ and $\boldsymbol{B}$ are, respectively, $N \times m$ and $q \times p$ full rank known matrices with $N > m$ and $p \ge q$, $\boldsymbol{\beta}$ is an $m \times q$ unknown coefficient matrix, and $\boldsymbol{e}$ is an $N \times p$ error matrix. Assume that the error $\boldsymbol{e}$ has elliptical density $|\Omega|^{-N/2}f(\mathrm{tr}\,\Omega^{-1}\boldsymbol{e}^t\boldsymbol{e})$ where Ω is a $p \times p$ unknown positive definite matrix, and $f(\cdot)$ is a nonnegative unknown function.

For providing a canonical form of (3.11), let $\Sigma = \Gamma\Omega\Gamma^t$ for orthogonal matrix $\Gamma = (\boldsymbol{B}^t(\boldsymbol{B}^t\boldsymbol{B})^{-1/2}, \boldsymbol{B}_0)$ with some matrix $\boldsymbol{B}_0$, and for $i, j = 1, 2$, Σ_{ij} is a matrix element of Σ with $q \times q$ matrix Σ_{11}. By making orthogonal transformations (Srivastava and Khatri(1979)), the density of $\boldsymbol{y}$ is written as

$$\begin{aligned}
|\Sigma|^{-N/2}f\Big(&\mathrm{tr}\,\Sigma_{11.2}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2\boldsymbol{\gamma})^t(\boldsymbol{x}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2\boldsymbol{\gamma}) + \mathrm{tr}\,\Sigma_{11.2}^{-1}\boldsymbol{v}_2^t\boldsymbol{v}_2 \\
&+\mathrm{tr}\,\Sigma_{11.2}^{-1}(\boldsymbol{v}_1 - (\boldsymbol{z}^t\boldsymbol{z})^{1/2}\boldsymbol{\gamma})^t(\boldsymbol{v}_1 - (\boldsymbol{z}^t\boldsymbol{z})^{1/2}\boldsymbol{\gamma}) \\
&+\mathrm{tr}\,\Sigma_{22}^{-1}\boldsymbol{x}_2^t\boldsymbol{x}_2 + \mathrm{tr}\,\Sigma_{22}^{-1}\boldsymbol{z}^t\boldsymbol{z}\Big)\, d\boldsymbol{x}_1 d\boldsymbol{x}_2 d\boldsymbol{v}_1 d\boldsymbol{v}_2 d\boldsymbol{z},
\end{aligned} \tag{3.12}$$

where $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{z}$ are, respectively, $m \times q$, $m \times (p-q)$, $(p-q) \times q$, $n \times q$ and $(N-m) \times (p-q)$ random matrices for $n = N - m - (p-q)$, and $\boldsymbol{\theta} = (\boldsymbol{A}\boldsymbol{A}^t)^{1/2}\boldsymbol{\beta}(\boldsymbol{B}\boldsymbol{B}^t)^{1/2}$, $\boldsymbol{\gamma} = \Sigma_{22}^{-1}\Sigma_{21}$. Denote $\boldsymbol{S}_{22} = \boldsymbol{z}^t\boldsymbol{z}$, $\boldsymbol{S}_{11.2} = \boldsymbol{v}_2^t\boldsymbol{v}_2$ and $\boldsymbol{S}_{21} = \boldsymbol{S}_{22}^{1/2}\boldsymbol{v}_1$.

The MLE of $\boldsymbol{\theta}$ in the normal distribution is given by

$$\hat{\boldsymbol{\theta}}^{ML} = \boldsymbol{x}_1 - \boldsymbol{x}_2\boldsymbol{S}_{22}^{-1/2}\boldsymbol{v}_1 = \left(I, -\boldsymbol{x}_2\boldsymbol{S}_{22}^{-1/2}\right)\begin{pmatrix}\boldsymbol{x}_1 \\ \boldsymbol{v}_1\end{pmatrix},$$

which is also MLE in the ECD model if $f(\cdot)$ is a decreasing function. For improving on $\hat{\boldsymbol{\theta}}^{ML}$, Kariya *et al.* (1996) considered the double shrinkage estimators

$$\hat{\boldsymbol{\theta}}(G_1, G_2) = \left(I, -\boldsymbol{x}_2 \boldsymbol{S}_{22}^{-1/2}\right) \begin{pmatrix} \boldsymbol{x}_1 + \boldsymbol{G}_1(\boldsymbol{x}_1, \boldsymbol{S}_{11.2}) \\ \boldsymbol{v}_1 + \boldsymbol{G}_2(\boldsymbol{v}_1, \boldsymbol{S}_{11.2} | \boldsymbol{x}_2, \boldsymbol{S}_{22}) \end{pmatrix}.$$

The risk matrix of $\hat{\boldsymbol{\theta}}(\boldsymbol{G}_1, \boldsymbol{G}_2)$ is written as

$$\begin{aligned} & \boldsymbol{R}_{G1}(\hat{\boldsymbol{\theta}}(\boldsymbol{G}_1, \boldsymbol{G}_2), (\boldsymbol{\theta}, \Sigma), f) \\ = \ & E\left[\left(\hat{\boldsymbol{\theta}}(\boldsymbol{G}_1, \boldsymbol{G}_2) - \boldsymbol{\theta}\right) \Sigma_{11.2}^{-1} \left(\hat{\boldsymbol{\theta}}(\boldsymbol{G}_1, \boldsymbol{G}_2) - \boldsymbol{\theta}\right)^t\right] \qquad (3.13) \\ = \ & E\left[\left(\boldsymbol{I}, -\boldsymbol{x}_2 \boldsymbol{S}_{22}^{-1/2}\right) \begin{pmatrix} \boldsymbol{x}_1 + \boldsymbol{G}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2 \boldsymbol{\gamma} \\ \boldsymbol{v}_1 + \boldsymbol{G}_2 - \boldsymbol{S}_{22}^{1/2} \boldsymbol{\gamma} \end{pmatrix} \Sigma_{11.2}^{-1} \right. \\ & \left. \times \left\{\left(I, -\boldsymbol{x}_2 \boldsymbol{S}_{22}^{-1/2}\right) \begin{pmatrix} \boldsymbol{x}_1 + \boldsymbol{G}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2 \boldsymbol{\gamma} \\ \boldsymbol{v}_1 + \boldsymbol{G}_2 - \boldsymbol{S}_{22}^{1/2} \boldsymbol{\gamma} \end{pmatrix}\right\}^t\right]. \end{aligned}$$

For the cross term,

$$\begin{aligned} & E\left[(\boldsymbol{x}_1 + \boldsymbol{G}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2\boldsymbol{\gamma}) \Sigma_{11.2}^{-1} (\boldsymbol{v}_1 + \boldsymbol{G}_2 - \boldsymbol{S}_{22}^{1/2}\boldsymbol{\gamma})^t \left(\boldsymbol{S}_{22}^{-1/2}\right)^t \boldsymbol{x}_2^t\right] \\ = \ & E\left[(\boldsymbol{x}_1 + \boldsymbol{G}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2\boldsymbol{\gamma}) \Sigma_{11.2}^{-1} (\boldsymbol{v}_1 - \boldsymbol{S}_{22}^{1/2}\boldsymbol{\gamma})^t \left(\boldsymbol{S}_{22}^{-1/2}\right)^t \boldsymbol{x}_2^t\right] \qquad (3.14) \\ & + E\left[(\boldsymbol{x}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2\boldsymbol{\gamma}) \Sigma_{11.2}^{-1} (\boldsymbol{v}_1 + \boldsymbol{G}_2 - \boldsymbol{S}_{22}^{1/2}\boldsymbol{\gamma})^t \left(\boldsymbol{S}_{22}^{-1/2}\right)^t \boldsymbol{x}_2^t\right] \\ & + E\left[(\boldsymbol{x}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2\boldsymbol{\gamma}) \Sigma_{11.2}^{-1} (\boldsymbol{v}_1 - \boldsymbol{S}_{22}^{1/2}\boldsymbol{\gamma})^t \left(\boldsymbol{S}_{22}^{-1/2}\right)^t \boldsymbol{x}_2^t\right] \\ & + E\left[\boldsymbol{G}_1 \Sigma_{11.2}^{-1} \boldsymbol{G}_2^t \left(\boldsymbol{S}_{22}^{-1/2}\right)^t \boldsymbol{x}_2^t\right]. \end{aligned}$$

Noting that the density function is symmetric at $\boldsymbol{x}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2\boldsymbol{\gamma} = 0$ and $\boldsymbol{v}_1 - \boldsymbol{S}_{22}^{1/2}\boldsymbol{\gamma} = \boldsymbol{0}$ gives that the first three terms of (3.14) are zero. If we assume the restriction on $\boldsymbol{G}_2$ as

$$\boldsymbol{G}_2(\cdot | -\boldsymbol{x}_2, \boldsymbol{S}_{22}) = \boldsymbol{G}_2(\cdot | \boldsymbol{x}_2, \boldsymbol{S}_{22}), \qquad (3.15)$$

then the same argument yields that the fourth term is equal to zero, so that the cross term is zero. Hence the risk matrix (3.13) is rewritten as

$$\begin{aligned} & \boldsymbol{R}_{G1}(\hat{\boldsymbol{\theta}}(\boldsymbol{G}_1, \boldsymbol{G}_2), (\boldsymbol{\theta}, \Sigma), f) \\ = \ & E\left[(\boldsymbol{x}_1 + \boldsymbol{G}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2\boldsymbol{\gamma}) \Sigma_{11.2}^{-1} (\boldsymbol{x}_1 + \boldsymbol{G}_1 - \boldsymbol{\theta} - \boldsymbol{x}_2\boldsymbol{\gamma})^t\right] \\ & + E\left[\boldsymbol{x}_2 \boldsymbol{S}_{22}^{-1/2} \left(\boldsymbol{v}_1 + \boldsymbol{G}_2 - \boldsymbol{S}_{22}^{1/2}\boldsymbol{\gamma}\right) \Sigma_{11.2}^{-1} (\boldsymbol{v}_1 + \boldsymbol{G}_2 - \boldsymbol{S}_{22}^{1/2}\boldsymbol{\gamma})^t \left(\boldsymbol{S}_{22}^{-1/2}\right)^t \boldsymbol{x}_2^t\right]. \end{aligned}$$

The above expression allows us to enjoy shrinking $\boldsymbol{x}_1$ and $\boldsymbol{v}_1$ doubly. Also the improvements on the MLE are reduced to those on $\boldsymbol{x}_1$ and $\boldsymbol{v}_1$ in multivariate regression models. The same arguments as in Section 3.2 are therefore used to establish the robust dominance results in the ECD model. For the scalar risk

$$R_{G2}(\hat{\boldsymbol{\theta}}(\boldsymbol{G}_1, \boldsymbol{G}_2), (\boldsymbol{\theta}, \Sigma), f) = \operatorname{tr} \boldsymbol{R}_{G1}(\hat{\boldsymbol{\theta}}(\boldsymbol{G}_1, \boldsymbol{G}_2), (\boldsymbol{\theta}, \Sigma), f),$$

the similar dominance results can be developed. Hence all improved procedures derived by Kariya *et al.* (1996) and Konno *et al.* (1994) are guaranteed to be robust in the ECD model.

For instance, the Efron-Morris type estimator $\hat{\boldsymbol{\theta}}^{EM} = \boldsymbol{x}_1 - \boldsymbol{x}_2\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21} - \boldsymbol{G}$ has the robust improvement on $\hat{\boldsymbol{\theta}}^{ML}$, where

$$\begin{aligned}
\boldsymbol{G} &= \left[c_1\boldsymbol{x}_1(\boldsymbol{x}_1^t\boldsymbol{x}_1)^{-1} - c_2\boldsymbol{x}_2(\boldsymbol{x}_2^t\boldsymbol{x}_2)^{-1}\boldsymbol{S}_{21}\left\{\boldsymbol{S}_{12}(\boldsymbol{x}_2^t\boldsymbol{x}_2)^{-1}\boldsymbol{S}_{21}\right\}^{-1}\right]\boldsymbol{S}_{11.2} \\
&\qquad\qquad\qquad\qquad \text{for } m \geq q+2 \text{ and } p-q \geq q+2, \\
&= c_1(\boldsymbol{x}_1\boldsymbol{S}_{11.2}^{-1}\boldsymbol{x}_1^t)^{-1}\boldsymbol{x}_1 - c_2\boldsymbol{x}_2(\boldsymbol{S}_{21}\boldsymbol{S}_{11.2}^{-1}\boldsymbol{S}_{12})^{-1}\boldsymbol{S}_{21} \\
&\qquad\qquad\qquad\qquad \text{for } q \geq m+2 \text{ and } q \geq p-q+2,
\end{aligned}$$

with

$$c_1 = \frac{m \vee q - m \wedge q - 1}{n + (2m-q)\wedge q + 1} \quad \text{and} \quad c_2 = \frac{q \vee (p-q) - q \wedge (p-q) - 1}{n + (2p-3q)\wedge q + 1},$$

for $a \vee b = \max(a,b)$ and $a \wedge b = \min(a,b)$.

For the noninvariant loss functions, an extension to the growth curve model was given by Tan(1991) in the case of a normal distribution. Combining the above arguments and the results of Section 2.2, we can provide robust dominance results for the noninvariant loss and enjoy the robust improvements of the double shrinkage estimators.

4 Robust Improvements in Estimation of the Covariance Matrix

The estimation of the covariance matrix of the normal distribution is one of interesting issues which have been addressed in lots of papers. Let $\boldsymbol{S}$ be a random matrix having a Wishart distribution with n degrees of freedom and the expectation $E[\boldsymbol{S}] = n\Sigma$. When estimator $\widehat{\Sigma}$ is evaluated through Stein's loss $\operatorname{tr}\widehat{\Sigma}\Sigma^{-1} - \log|\widehat{\Sigma}\Sigma^{-1}| - p$, James and Stein(1961) showed that unbiased estimator $\widehat{\Sigma}^{UB} = n^{-1}\boldsymbol{S}$ is improved on by

$$\widehat{\Sigma}^{JS} = \boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^t,$$

where $\boldsymbol{D} = \operatorname{diag}(d_1,\ldots,d_p)$ for $d_i = 1/(n+p+1-2i)$, $i = 1,\ldots,p$ and $\boldsymbol{T}$ is a $p \times p$ lower triangular matrix such that $\boldsymbol{S} = \boldsymbol{T}\boldsymbol{T}^t$. It is also known that $\widehat{\Sigma}^{JS}$ is further dominated by Stein's orthogonally invariant estimator

$$\widehat{\Sigma}^{ST} = \boldsymbol{H}\operatorname{diag}(d_1\ell_1,\ldots,d_p\ell_p)\boldsymbol{H}^t,$$

where $\boldsymbol{H}$ is a $p \times p$ orthogonal matrix and $\ell_1,\ldots,\ell_p$ are eigen values of $\boldsymbol{S}$ such that $\boldsymbol{S} = \boldsymbol{H}\operatorname{diag}(\ell_1,\ldots,\ell_p)\boldsymbol{H}^t$ and $\ell_1 \geq \cdots \geq \ell_p$.

The purpose of this section is to investigate whether these decision-theoretic results hold still for every ECD model.

Theorem 4.1 *For the estimation of Σ in the canonical form (2.1), the James-Stein estimator $\widehat{\Sigma}^{JS}$ is better than $\widehat{\Sigma}^{UB}$ uniformly for every unknown function $f(\cdot)$.*

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

$$\begin{aligned}
\Delta_{C1} &= R(\widehat{\Sigma}^{UB},(\boldsymbol{\theta},\Sigma),f) - R(\widehat{\Sigma}^{JS},(\boldsymbol{\theta},\Sigma),f) \\
&= E^f_{\theta,\Sigma}\left[n^{-1}\operatorname{tr}\boldsymbol{S}\Sigma^{-1} - \log|n^{-1}\boldsymbol{S}\Sigma^{-1}| - \operatorname{tr}\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^t\Sigma^{-1} + \log|\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^t\Sigma^{-1}|\right] \\
&= E^f_{\xi,I}\left[n^{-1}\operatorname{tr}\boldsymbol{S} + p\log n - \operatorname{tr}\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^t + \sum_{i=1}^{p}\log d_i\right]. \qquad (4.1)
\end{aligned}$$

Note that

$$E^f_{\xi,I}\left[n^{-1}\mathrm{tr}\,\boldsymbol{S}\right] = E^f_{\xi,I}\left[n^{-1}\sum_{i=1}^{p}\boldsymbol{z}_i^t\boldsymbol{z}_i\right] = p\times E^F_{\xi,I}[1] \tag{4.2}$$

for $\boldsymbol{z} = (\boldsymbol{z}_1,\ldots,\boldsymbol{z}_p)$. If we can show that

$$E^f_{\xi,I}[\boldsymbol{T}^t\boldsymbol{T}] = \boldsymbol{D}^{-1}E^F_{\xi,I}[1], \tag{4.3}$$

then combining (4.1), (4.2) and (4.3) gives

$$\Delta_{C1} = E^f_{\xi,I}\left[p\log n - \sum_{i=1}^{p}\log\,(n+p+1-2i)\right],$$

which is nonnegative as checked easily.

We shall now verify the condition (4.3) to complete the proof. For the purpose, $\boldsymbol{S}$ and $\boldsymbol{T}$ are decomposed by $\boldsymbol{S} = (\boldsymbol{S}_{ij})$ and $\boldsymbol{T} = (\boldsymbol{T}_{ij})$ for $i,j = 1,2$ with scalars S_{22}, T_{22} and $\boldsymbol{T}_{12} = \boldsymbol{0}$. Since $\boldsymbol{S}_{11} = \boldsymbol{T}_{11}\boldsymbol{T}_{11}^t$, $\boldsymbol{S}_{12} = \boldsymbol{T}_{11}\boldsymbol{T}_{21}^t$ and $S_{22} = \boldsymbol{T}_{21}\boldsymbol{T}_{21}^t + T_{22}^2$, we observe that

$$\begin{aligned}
(\boldsymbol{T}^t\boldsymbol{T})_{11} &= \boldsymbol{T}_{11}^t\boldsymbol{T}_{11} + \boldsymbol{T}_{21}^t\boldsymbol{T}_{21} = \boldsymbol{T}_{11}^t\boldsymbol{T}_{11} + \boldsymbol{T}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{S}_{12}^t(\boldsymbol{T}_{11}^t)^{-1},\\
(\boldsymbol{T}^t\boldsymbol{T})_{12} &= \boldsymbol{T}_{21}^t T_{22} = \boldsymbol{T}_{11}^{-1}\boldsymbol{S}_{12}\sqrt{S_{22} - \boldsymbol{S}_{12}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}},\\
(\boldsymbol{T}^t\boldsymbol{T})_{22} &= T_{22}^2 = S_{22} - \boldsymbol{S}_{12}^t\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}.
\end{aligned}$$

Let $\boldsymbol{S}_{ij} = \boldsymbol{z}_i^t\boldsymbol{z}_j$ for $\boldsymbol{z} = (\boldsymbol{z}_1,\boldsymbol{z}_2)$ with $n\times 1$ vector $\boldsymbol{z}_2$, and let $(\boldsymbol{v}_1^t,\boldsymbol{v}_2^t)^t = \boldsymbol{H}\boldsymbol{z}_2$ with $(p-1)\times 1$ vector $\boldsymbol{v}_1$ for $n\times n$ orthogonal matrix $\boldsymbol{H}$ such that $(\boldsymbol{H}\boldsymbol{z}_1)^t = (\boldsymbol{T}_{11},\boldsymbol{0})$. Then we have $\boldsymbol{S}_{12} = \boldsymbol{z}_1^t\boldsymbol{z}_2 = \boldsymbol{T}_{11}\boldsymbol{v}_1$, so that

$$\begin{aligned}
E^f_{\xi,I}\left[\boldsymbol{T}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{S}_{12}^t(\boldsymbol{T}_{11}^t)^{-1}\right] &= E^f_{\xi,I}\left[\boldsymbol{v}_1\boldsymbol{v}_1^t\right] = \boldsymbol{I}_{p-1}E^F_{\xi,I}[1], && (4.4)\\
E^f_{\xi,I}\left[S_{22} - \boldsymbol{S}_{12}^t\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}\right] &= E^f_{\xi,I}\left[\boldsymbol{z}_2^t(\boldsymbol{I}_n - \boldsymbol{z}_1(\boldsymbol{z}_1^t\boldsymbol{z}_1)^{-1}\boldsymbol{z}_1^t)\boldsymbol{z}_2\right]\\
&= E^f_{\xi,I}[\boldsymbol{v}_2^t\boldsymbol{v}_2] = (n-p+1)E^F_{\xi,I}[1], && (4.5)\\
E^f_{\xi,I}\left[\boldsymbol{T}_{11}^{-1}\boldsymbol{S}_{12}\sqrt{S_{22} - \boldsymbol{S}_{12}^t\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}}\right] &= E^f_{\xi,I}\left[\boldsymbol{v}_1\sqrt{\boldsymbol{v}_2^t\boldsymbol{v}_2}\right] = 0. && (4.6)
\end{aligned}$$

On the basis of (4.4), (4.5) and (4.6), the equation (4.3) is verified by the induction. For $p=2$, noting that $E^f_{\xi,I}[\boldsymbol{T}_{11}^t\boldsymbol{T}_{11}] = E^f_{\xi,I}[\boldsymbol{S}_{11}] = nE^F_{\xi,I}[1]$, we can easily see that $E^f_{\xi,I}[\boldsymbol{T}^t\boldsymbol{T}] = \mathrm{diag}\,(n+1, n-1)E^F_{\xi,I}[1]$. For $p\geq 3$, suppose that $E^f_{\xi,I}[\boldsymbol{T}_{11}^t\boldsymbol{T}_{11}] = \mathrm{diag}\,(n+(p-1)+1-2i, i=1,\ldots,p-1)E^F_{\xi,I}[1]$. Then from (4.4), $E^f_{\xi,I}[\boldsymbol{T}_{11}^t\boldsymbol{T}_{11} + \boldsymbol{T}_{21}^t\boldsymbol{T}_{21}] = \mathrm{diag}\,(n+p+1-2i, i=1,\ldots,p-1)E^F_{\xi,I}[1]$. Hence from (4.5) and (4.6), we get (4.3) and the proof is complete. △

For the assertion of the robustness of Stein's result, the following lemma is essential.

Lemma 4.1. *Let* $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t$, $\boldsymbol{L} = \mathrm{diag}\,(\ell_1,\ldots,\ell_p)$, $\ell_1\geq\ldots\geq\ell_p$, *and consider the estimator* $\widehat{\Sigma}(\boldsymbol{\phi}) = \boldsymbol{H}\mathrm{diag}\,(\phi_1(\boldsymbol{L}),\ldots,\phi_p(\boldsymbol{L}))\,\boldsymbol{H}^t$. *Then under suitable conditions corresponding to those of Lemma 2.2,*

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

This lemma is immediately derived from Lemma 2.2 and the equation

$$\mathrm{tr}\,[\boldsymbol{D}_S\widehat{\Sigma}(\boldsymbol{\phi})] = \sum_{i\neq j}\phi_i(\boldsymbol{L})/(\ell_i-\ell_j) + \sum_i\partial\phi_i(\boldsymbol{L})/\partial\ell_i$$

as evaluated by Dey and Srinivasan(1985).

Theorem 4.2. *$\widehat{\Sigma}^{ST}$ is better than $\widehat{\Sigma}^{JS}$ uniformly for every unknown $f(\cdot)$.*

Proof. Using (4.3) and Lemma 4.1, we can write the risk difference of estimators $\widehat{\Sigma}^{JS}$ and $\widehat{\Sigma}^{ST}$ as

$$\begin{aligned} \Delta_{C2} &= R(\widehat{\Sigma}^{JS},(\boldsymbol{\theta},\Sigma),f) - R(\widehat{\Sigma}^{ST},(\boldsymbol{\theta},\Sigma),f) \\ &= E^f_{\theta,\Sigma}\left[\mathrm{tr}\,\boldsymbol{D}\boldsymbol{T}^t\Sigma^{-1}\boldsymbol{T}\right] - E^f_{\theta,\Sigma}\left[\mathrm{tr}\,\boldsymbol{H}\mathrm{diag}\,(d_1\ell_1,\ldots,d_p\ell_p)\boldsymbol{H}^t\Sigma^{-1}\right] \qquad (4.7) \\ &= E^F_{\theta,\Sigma}[p] - E^F_{\theta,\Sigma}\left[2\sum_{i>j}\frac{d_i\ell_i - d_j\ell_j}{\ell_i-\ell_j} + 2\sum_i d_i + (n-p-1)\sum_i d_i\right]. \end{aligned}$$

Using the equation

$$\frac{d_i\ell_i - d_j\ell_j}{\ell_i-\ell_j} = \frac{d_i-d_j}{\ell_i-\ell_j}\ell_i + d_j,$$

we can rewrite Δ_{C2} as

$$\begin{aligned} \Delta_{C2} &= -E^F_{\theta,\Sigma}\left[2\sum_{i>j}\frac{d_i-d_j}{\ell_i-\ell_j}\ell_i + \sum_i(n+p+1-2i)d_i - p\right] \\ &= -E^F_{\theta,\Sigma}\left[2\sum_{i>j}\frac{d_i-d_j}{\ell_i-\ell_j}\ell_i\right], \end{aligned}$$

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 $\ell_i < \ell_j$, so that we get that $\Delta_{C2} \geq 0$, and the proof is complete. $\triangle$

Two major dominance results in estimation of the covariance matrix have been thus established to be robust in our sense. Another orthogonally invariant estimator of the form

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

was proposed by Takemura(1984) where $O(p)$ designates a class of $p\times p$ orthogonal matrices and $\boldsymbol{T}_\Gamma$ is a $p\times p$ lower triangular matrix such that $\Gamma^t\boldsymbol{S}\Gamma = \boldsymbol{T}_\Gamma\boldsymbol{T}_\Gamma^t$. It can be seen that this estimator is superior to Σ^{JS} uniformly for every function $f(\cdot)$. Also Sheena and Takemura(1992) proved that non-order-preserving estimators are improved on by the corresponding order-preserving estimators. In other words, let $\widehat{\Sigma}(\boldsymbol{\phi}) = \boldsymbol{H}\mathrm{diag}\,(\phi_1(\boldsymbol{L}),\ldots,\phi_p(\boldsymbol{L}))\boldsymbol{H}^t$ be an orthogonally invariant estimator and let $\widehat{\Sigma}(\boldsymbol{\phi}^O)$ be the order-preserving estimator given by modifying $\widehat{\Sigma}(\boldsymbol{\phi})$ as $\widehat{\Sigma}(\boldsymbol{\phi}^O) = \boldsymbol{H}\mathrm{diag}\,(\phi_1^O(\boldsymbol{L}),\ldots,\phi_p^O(\boldsymbol{L}))\boldsymbol{H}^t$, where $\phi_i^O(\boldsymbol{L})$ is the ith largest element in $(\phi_1(\boldsymbol{L}),\ldots,\phi_p(\boldsymbol{L}))$, that is, $\phi_1^O(\boldsymbol{L}) \geq \cdots \geq \phi_p^O(\boldsymbol{L})$. Then $\widehat{\Sigma}(\boldsymbol{\phi}^O)$ is better than $\widehat{\Sigma}(\boldsymbol{\phi})$ in the normal distribution if $P_\Sigma(\phi_i^O(\boldsymbol{L}) \neq \phi_i(\boldsymbol{L})$ for some $i) > 0$ for some Σ. The robustness of this dominance can be guaranteed 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 also demonstrates the inadmissibility of $\widehat{\Sigma}^{ST}$ for $p\geq 2$ and every nonincreasing function $f(\cdot)$.

In the ECD model, $n^{-1}\boldsymbol{S}$ is an unbiased estimator of $\Sigma^* = E^f_{\theta,\Sigma}[n^{-1}\boldsymbol{S}] = E^F_{\xi,I}[1]\Sigma$. 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 Σ^*.

5 Robust Improvements in Other Estimation Problems

5.1 Estimation of variance with unknown locations

We shall, in this section, investigate whether the robust improvements can be asserted in other estimation issues. The robustness of improvements shown in the previous sections is technically grounded on the Stein and the Haff identities given by Lemmas 2.1 and 2.2. The dominance results in some estimation problems have, on the other hand, been established without using their identities. In the estimation of a variance of a normal distribution with an unknown mean, for instance, the dominance result is proved through an infinite series expression of a noncentral chi square distribution. It is thus of great interest to argue whether the robustness can be extended to such a situation of the dominance.

We here focus on the problem of estimating the scale parameter σ^2 in the model (1.1) with the density (1.2) and $\sigma^2 = \Sigma$ for $p = 1$. The canonical form (2.1) is written as

$$\sigma^{-N} f(\sigma^{-1}||\boldsymbol{x} - \boldsymbol{\theta}||^2 + \sigma^{-1}||\boldsymbol{z}||^2), \tag{5.1}$$

where $\boldsymbol{x} \in \boldsymbol{R}^m$, $\boldsymbol{\theta} \in \boldsymbol{R}^m$, $\boldsymbol{z} \in \boldsymbol{R}^{N-m}$ and $||\boldsymbol{z}||^2 = \boldsymbol{z}^t\boldsymbol{z}$. Letting $n = N - m$, Stein(1964) showed in the normal distribution that unbiased estimator $\hat{\sigma}^{2UB} = n^{-1}||\boldsymbol{z}||^2$ of σ^2 is dominated by

$$\hat{\sigma}^{2ST} = \min\{n^{-1}||\boldsymbol{z}||^2,\ (n+m)^{-1}(||\boldsymbol{z}||^2 + ||\boldsymbol{x}||^2)\}$$

relative to the loss

$$L(\hat{\sigma}^2, \sigma^2) = \hat{\sigma}^2/\sigma^2 - \log\hat{\sigma}^2/\sigma^2 - 1, \tag{5.2}$$

which can be derived from the Kullback-Leibler distance. Our purpose is to investigate the robustness of the dominance result.

Let $\boldsymbol{u} = (u_1, \ldots, u_m)^t = \sigma^{-1}\boldsymbol{H}\boldsymbol{x}$, $\mu = (\sigma^{-1}||\boldsymbol{\theta}||)^2$ and $\boldsymbol{v} = \sigma^{-1}\boldsymbol{z}$ for $m \times m$ orthogonal matrix $\boldsymbol{H}$ such that $\boldsymbol{H\theta} = (||\boldsymbol{\theta}||, 0, \ldots, 0)^t$. The joint density of $\boldsymbol{u}$ and $\boldsymbol{v}$ is rewritten by $f(||\boldsymbol{u}||^2 - 2\sqrt{\mu}u_1 + \mu + ||\boldsymbol{v}||^2)$, and for $T = ||\boldsymbol{u}||^2$ and $S = ||\boldsymbol{v}||^2$, we begin with obtaining the joint density of (T, S). The following formula is quite useful for our purpose. For any function $h(\cdot)$ and $\boldsymbol{y} \in \boldsymbol{R}^k$,

$$\int\int h(||\boldsymbol{y}||^2) I(s = ||\boldsymbol{y}||^2) d\boldsymbol{y} ds = c_k \int s^{k/2-1} h(s) ds \tag{5.3}$$

where $c_k = \{\Gamma(k/2)\}^{-1}\pi^{k/2}$. This formula can be verified by using the transformation to the polar coordinates (see Takemura(1991)). It is also a special case of Corollary 3.2.3 of Srivastava and Khatri(1979).

Lemma 5.1. *Suppose that the function $f(\cdot)$ belongs to the $C^{(\infty)}$ class, namely being infinite-times continuously differentiable, and permits the Taylor expansion. Then the joint density of (T, S) is given by*

$$g(t, s; \mu) = \sum_{k=0}^{\infty} \frac{(\mu/\pi)^k}{k!} c_{m+2k} c_n t^{(m+2k)/2-1} s^{n/2-1} f^{(2k)}(t + s + \mu), \tag{5.4}$$

where $f^{(2k)}(x) = (d^{2k}/dx^{2k}) f(x)$.

Proof. Expanding $f(||\boldsymbol{u}||^2 - 2\sqrt{\mu}u_1 + \mu + ||\boldsymbol{v}||^2)$ with respect to $-2\sqrt{\mu}u_1$ gives the infinite series

$$\sum_{k=0}^{\infty} \frac{(-2\sqrt{\mu}u_1)^k}{k!} f^{(k)}(||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 + \mu),$$

so that

$$
\begin{aligned}
1 &= \int \cdots \int f(||\boldsymbol{u}||^2 - 2\sqrt{\mu}u_1 + \mu + ||\boldsymbol{v}||^2) du_1 \prod_{i=2}^{m} du_i d\boldsymbol{v} \\
&= \int \cdots \int \sum_{k=0}^{\infty} \frac{(4\mu)^k}{(2k)!} u_1^{2k} f^{(2k)}(||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 + \mu) du_1 \prod_{i=2}^{m} du_i d\boldsymbol{v}. \qquad (5.5)
\end{aligned}
$$

By using the formula (5.3) several times, the r.h.s. of (5.5) is expressed as

$$
\begin{aligned}
&\int \cdots \int \sum_{k=0}^{\infty} \frac{(4\mu)^k}{(2k)!} c_1 x^k \cdot x^{1/2-1} f^{(2k)}(x + \sum_{i=2}^{m} u_i^2 + ||\boldsymbol{v}||^2 + \mu) dx \prod_{i=2}^{m} du_i d\boldsymbol{v} \\
&= \int\int \sum_{k=0}^{\infty} \frac{(4\mu)^k}{(2k)!} \frac{c_1}{c_{2k+1}} f^{(2k)}(||\boldsymbol{u}^*||^2 + ||\boldsymbol{v}||^2 + \mu) d\boldsymbol{u}^* d\boldsymbol{v} \\
&= \int\int \sum_{k=0}^{\infty} \frac{(4\mu)^k}{(2k)!} \frac{c_1}{c_{2k+1}} c_{m+2k} c_m t^{(m+2k)/2-1} s^{n/2-1} f^{(2k)}(t + s + \mu) dt ds, \qquad (5.6)
\end{aligned}
$$

where $\boldsymbol{u}^* \in \boldsymbol{R}^{m+2k}$. Noting that $\{4^k/(2k)!\}c_1/c_{2k+1} = (k!\pi^k)^{-1}$, we get the expression (5.4). △

On the basis of Lemma 5.1, we get a result concerning the improvement on $\hat{\sigma}^{2UB}$. Letting $W = T + S$ and $Z = T/(T + S)$, we see that the conditional expectation given Z in the risk of the estimator $(||\boldsymbol{x}||^2 + ||\boldsymbol{z}||^2)\phi(||\boldsymbol{x}||^2/(||\boldsymbol{x}||^2 + ||\boldsymbol{z}||^2))$, being rewritten by $\sigma^2 W\phi(Z)$, is minimized at $\phi(Z) = \{E[W|Z]\}^{-1}$. From Lemma 5.1, the joint density of (W, Z) is given by

$$
h(w, z; \mu) = \sum_{k=0}^{\infty} \frac{(\mu/\pi)^k}{k!} c_{m+2k} c_n \, z^{(m+2k)/2-1}(1 - z)^{n/2-1} w^{(m+n+2k)/2-1} f^{(2k)}(w + \mu),
$$

which provides that $\{E[W|Z]\}^{-1} \leq A_{m+n}(f)$, where

$$
A_{m+n}(f) = \sup_{k\geq 0, \mu\geq 0} \frac{\int w^{(m+n+2k)/2-1} f^{(2k)}(w + \mu) dw}{\int w^{(m+n+2k)/2} f^{(2k)}(w + \mu) dw}.
$$

Hence we get

Proposition 5.1. *If there exists A^* such that $A_{m+n}(f) \leq A^* < n^{-1}$, then $\hat{\sigma}^{2UB} = n^{-1}||\boldsymbol{z}||^2$ is improved on by*

$$
\hat{\sigma}^{2ST}(A^*) = \min\{n^{-1}||\boldsymbol{z}||^2, A^*(||\boldsymbol{x}||^2 + ||\boldsymbol{z}||^2)\}
$$

relative to the loss (5.2).

It is noted that the constant $A_{m+n}(f)$ generally depends on the function $f(\cdot)$. However, by imposing restrictions on the class of the distributions, we can get improved estimators independent of f. One of the restrictions we treat is to assume that

$$
e^{w/2} f^{(2k)}(w + \mu) \text{ is nondecreasing in } w \text{ for every } k \geq 0. \qquad (5.7)
$$

Noting that $1/w$ and $e^{w/2} f^{(2k)}(w+\mu)$ are monotone in opposite directions under the assumption (5.7), we get the inequality

$$
\frac{\int w^{(m+n+2k)/2-1} f^{(2k)}(w + \mu) dw}{\int w^{(m+n+2k)/2} f^{(2k)}(w + \mu) dw} \leq \frac{\int w^{(m+n+2k)/2-1} e^{-w/2} dw}{\int w^{(m+n+2k)/2} e^{-w/2} dw}, \qquad (5.8)
$$

and the r.h.s. of (5.8) is equal to $1/(m+n+2k)$, so that $A_{m+n}(f) = 1/(m+n)$.

For instance, consider a class of contaminated (or mixture) normal distributions such that the error term $\boldsymbol{e}$ in the model (1.1) has

$$(1-\lambda)\mathcal{N}_N(\mathbf{0}, \sigma^2 \boldsymbol{I}_N) + \lambda \mathcal{N}_N(\mathbf{0}, \sigma^{*2} \boldsymbol{I}_N) \tag{5.9}$$

where λ, σ^2 and σ^{*2} are unknown parameters satisfying $\tau = \sigma^{*2}/\sigma^2 \geq 1$. This model means some data with a larger variance can be taken with probability λ. The function $f(w+\mu)$ is thus represented by

$$f(w+\mu) = (1-\lambda)(2\pi)^{-N/2}\exp\left\{-\frac{w+\mu}{2}\right\} + \lambda(2\pi\tau)^{-N/2}\exp\left\{-\frac{w+\mu}{2\tau}\right\}, \tag{5.10}$$

and it is easy to check the assumption (5.7). Hence the robustness of the Stein's dominance result still holds within the model (5.9).

One choice of distributions with heavier tails is a multivariate t-distribution whose density is given by

$$\frac{\nu^{\nu/2}\Gamma((N+\nu)/2)}{\pi^{N/2}\sigma^N\Gamma(\nu/2)}(\nu + ||\mathbf{e}||^2/\sigma^2)^{-(N+\nu)/2}, \quad \nu \geq 1, \tag{5.11}$$

which approaches a normal density as ν tends to infinity. The function $f^{(2k)}(w+\mu) = (const.)(w+\mu+\nu)^{-(N+\nu)/2-2k}$, unfortunately, does not satisfy the assumption (5.7). Thereby, $A_{m+n}(f)$ is directly calculated as

$$\begin{aligned} A_{m+n}(f) &= \sup_{k\geq 0, \mu \geq 0} \frac{\int_0^\infty w^{(m+n)/2+k-1}(w+\mu+\nu)^{-(N+\nu)/2-2k}dw}{\int_0^\infty w^{(m+n)/2+k}(w+\mu+\nu)^{-(N+\nu)/2-2k}dw} \\ &= \sup_{k\geq 0, \mu\geq 0}\left\{\frac{\nu-2+2k}{(\mu+\nu)(n+m+2k)}\right\}, \end{aligned} \tag{5.12}$$

which is smaller than or equal to $(n+m)^{-1}$ if $\nu \geq n+m+2$, or ν^{-1} if $2 < \nu < n+m+2$. When the class of the multivariate t-distributions is restricted to the case $\nu \geq \ell$, we denote the restricted class by $\mathcal{F}_t(\ell)$. From Proposition 5.1 and the above arguments, therefore, we see that $\hat{\sigma}^{2ST}$ is better than $\hat{\sigma}^{2UB}$ for every $f \in \mathcal{F}_t(m+n+2)$, that is, the improvement is robust in the neighborhood $\mathcal{F}_t(m+n+2)$ of a normal distribution. From (5.12), it is also noted that $A_{n+m}(f) \leq (n+1)^{-1}$ for $\nu \geq n+1$ and $n \geq 2$. Hence from Proposition 5.1, $\hat{\sigma}^{2UB}$ is improved on by $\hat{\sigma}^{2ST}((n+1)^{-1})$ for every f in larger class $\mathcal{F}_t(n+1)$ although the risk gain will be quite small.

When the estimation problem of σ^2 of the multivariate t-distribution is considered for fixed ν, apart from the contents of the robust improvements, the above arguments provide the dominance result that unbiased estimator $\hat{\sigma}^{2UB}(\nu) = \nu^{-1}(\nu-2)n^{-1}||\boldsymbol{z}||^2$ of σ^2 is improved on by estimators

$$\begin{aligned} &\nu^{-1}(\nu-2)\min\{n^{-1}||\boldsymbol{z}||^2,\ (n+m)^{-1}(||\boldsymbol{x}||^2 + ||\boldsymbol{z}||^2)\} && \text{if } \nu \geq n+m+2, \\ &\nu^{-1}(\nu-2)\min\{n^{-1}||\boldsymbol{z}||^2,\ (\nu-2)^{-1}(||\boldsymbol{x}||^2 + ||\boldsymbol{z}||^2)\} && \text{if } n+2 < \nu < n+m+2. \end{aligned}$$

5.2 Estimation of ordered parameters in a mixed linear model

Let us consider the linear regression model (1.1) with $m = p = 1$, $\boldsymbol{A} = \boldsymbol{j}_N = (1, \ldots, 1)^t \in \boldsymbol{R}^N$, $\boldsymbol{\beta} = \mu \in \boldsymbol{R}$ and a covariance structure, that is, $\boldsymbol{y} = \mu \boldsymbol{j}_N + \boldsymbol{e}$, the $N \times 1$ error vector $\boldsymbol{e}$ having density

$$|\Omega|^{-1/2} f\left(\boldsymbol{e}^t \Omega^{-1} \boldsymbol{e}\right), \tag{5.13}$$

where Ω is a $N \times N$ matrix with the covariance structure

$$\Omega = \sigma^2 \boldsymbol{I}_N + \sigma_a^2 \oplus_{i=1}^k \boldsymbol{J}_i \tag{5.14}$$

for $\oplus_{i=1}^k \boldsymbol{J}_i = \mathrm{diag}(\boldsymbol{J}_1, \ldots, \boldsymbol{J}_k)$, $\boldsymbol{J}_i = \boldsymbol{j}_r \boldsymbol{j}_r^t$ and $N = kr$. In the case of a normal distribution, this corresponds to a one-way mixed linear model with two variance components:

$$\boldsymbol{e} = \boldsymbol{\alpha} \otimes \boldsymbol{j}_r + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\varepsilon}$ are independent random variables with $\boldsymbol{\alpha} \sim \mathcal{N}_k(\boldsymbol{0}, \sigma_a^2 \boldsymbol{I}_k)$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}_N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$.

For providing a canonical form, consider $1 \times N$ vector $\boldsymbol{H}_1 = N^{-1/2} \boldsymbol{j}_N^t$ and $(N-1) \times N$ matrix $\boldsymbol{H}_2$ such that $\boldsymbol{H}_2 \boldsymbol{j}_N = \boldsymbol{0}$ and $\boldsymbol{H}_2 \boldsymbol{H}_2^t = \boldsymbol{I}_{N-1}$. Let $x = \boldsymbol{H}_1 \boldsymbol{y}$ and $\boldsymbol{z} = \boldsymbol{H}_2 \boldsymbol{y}$. The joint density of $(x, \boldsymbol{z})$ is written by

$$|\Omega|^{-1/2} f\left(D(x) + \boldsymbol{z}^t (\sigma^2 \boldsymbol{I}_{N-1} + \sigma_a^2 \boldsymbol{H}_2 \{\oplus_{i=1}^k \boldsymbol{J}_i\} \boldsymbol{H}_2^t)^{-1} \boldsymbol{z}\right),$$

where $D(x) = (x - \mu\sqrt{N})^2 / (\sigma^2 + r\sigma_a^2)$. It can be easily seen that $\boldsymbol{H}_2 \{\oplus_{i=1}^k \boldsymbol{J}_i\} \boldsymbol{H}_2^t = r\boldsymbol{E}_2$ where $\boldsymbol{E}_2$ is an $(N-1) \times (N-1)$ idempotent matrix with $\mathrm{rank}(\boldsymbol{E}_2) = k-1$. Letting $\boldsymbol{E}_1 = \boldsymbol{I}_{N-1} - \boldsymbol{E}_2$ with $\mathrm{rank}(\boldsymbol{E}_1) = N - k$, we see that

$$\sigma^2 \boldsymbol{I}_{N-1} + \sigma_a^2 \boldsymbol{H}_2 \{\oplus_{i=1}^k \boldsymbol{J}_i\} \boldsymbol{H}_2^t = \sigma^2 \boldsymbol{E}_1 + (\sigma^2 + r\sigma_a^2) \boldsymbol{E}_2.$$

Let $S = \boldsymbol{y}^t \boldsymbol{E}_1 \boldsymbol{y}$ and $S^* = \boldsymbol{y}^t \boldsymbol{E}_2 \boldsymbol{y}$, and use the simple notations $n = N - k$, $n^* = k - 1$ and $\sigma^{*2} = \sigma^2 + r\sigma_a^2$. Applying the formula (5.3) and integrating out with respect to x, we get a joint density of (S, S^*) given by

$$c_n c_{n^*} (\sigma^2)^{-n/2} (\sigma^{*2})^{-n^*/2} s^{n/2-1} (s^*)^{n^*/2-1} g\left(\sigma^{-2} s + (\sigma^{*2})^{-1} s^*\right), \tag{5.15}$$

where

$$g(z) = \int (\sigma^*)^{-1} f(D(x) + z) dx. \tag{5.16}$$

It is noted that the parameters σ^2 and σ^{*2} possess the order relationship $\sigma^2 \le \sigma^{*2}$. The estimation issues of these ordered parameters σ^2 and σ^{*2} are considered here where the estimators are evaluated based on the same types of loss functions as in (5.2). In the normal distribution, the unbiased (or best scale equivariant) estimators of σ^2 and σ^{*2} are given by $\hat{\sigma}_{UB}^2 = n^{-1} S$ and $\hat{\sigma}_{UB}^2 = n^{*-1} S^*$, which are dominated by their truncation rules

$$\hat{\sigma}_{TR}^2 = \min\{n^{-1} S,\ (n + n^*)^{-1} (S + S^*)\}, \tag{5.17}$$
$$\hat{\sigma}_{TR}^{*2} = \max\{n^{*-1} S^*,\ (n + n^*)^{-1} (S + S^*)\}. \tag{5.18}$$

Our interest is to investigate whether these dominance results maintain the robustness in the ECD model.

For estimation of σ^2, more generally, we treat the estimators of the form $\hat{\sigma}^2(\psi) = S\psi(S^*/S)$ and obtain the condition on $\psi(\cdot)$ for $\hat{\sigma}^2(\psi)$ dominating $\hat{\sigma}_{UB}^2$ in terms of the risk

$$R_1(\omega; S\psi\left(\frac{S^*}{S}\right)) = E\left[\frac{S}{\sigma^2}\psi\left(\frac{S^*}{S}\right) - \log\frac{S}{\sigma^2}\psi\left(\frac{S^*}{S}\right) - 1\right].$$

Proposition 5.2. *Assume that*

(a) $\psi(w)$ is nondecreasing and $\lim_{w\to\infty}\psi(w) = a_0$,

(b) $\psi(w) \geq \psi_0(w) B_{n+n^*}(g)$ *where for* $g(z)$ *defined by (5.16),*

$$\psi_0(w) = \frac{\int_0^w x^{n^*/2-1}(1+x)^{-(n+n^*)/2}dx}{\int_0^w x^{n^*/2-1}(1+x)^{-(n+n^*)/2-1}dx} \tag{5.19}$$

and

$$B_{n+n^*}(g) = \frac{\int_0^\infty y^{(n+n^*)/2-1} g(y)dy}{\int_0^\infty y^{(n+n^*)/2} g(y)dy}. \tag{5.20}$$

Then $R_1(\omega; S_1\psi(S^*/S)) \leq R_1(\omega; a_0 S)$ *uniformly for every* ω.

Proof. Since $\lim_{w\to\infty}\psi(w) = n^{-1}$, the *Integral-Expression-of-Risk-Difference* (IERD) method used in Takeuchi(1991), Kubokawa(1994, 1995) and Kubokawa and Srivastava(1996) can be applied to get the following equations:

$$\begin{aligned} & R_1(\omega; a_0 S) - R_1\left(\omega; S\psi\left(\frac{S^*}{S}\right)\right) \\ &= E\left[\left\{\frac{S}{\sigma^2}\psi\left(\frac{S^*}{S}t\right) - \log\frac{S}{\sigma^2}\psi\left(\frac{S^*}{S}t\right) - 1\right\}\Big|_{t=1}^{\infty}\right] \\ &= E\left[\int_1^\infty \frac{d}{dt}\left\{\frac{S}{\sigma^2}\psi\left(\frac{S^*}{S}t\right) - \log\frac{S}{\sigma^2}\psi\left(\frac{S^*}{S}t\right) - 1\right\}dt\right]. \end{aligned} \tag{5.21}$$

Let $v = S/\sigma^2$ and $u = S^*/\sigma^{*2}$ and from (5.15), the joint density of (v, u) is given by

$$h(v,u) = c_n c_{n^*} v^{n/2-1} u^{n^*/2-1} g(v+u). \tag{5.22}$$

Carrying out the differentiation in (5.21) gives

$$\begin{aligned} & E\left[\int_1^\infty \left\{\frac{S}{\sigma^2} - \frac{1}{\psi(S^*t/S)}\right\}\frac{S^*}{S}\psi'\left(\frac{S^*}{S}t\right)dt\right] \\ &= \int\int\int_1^\infty \left\{v - \frac{1}{\psi(\theta ut/v)}\right\}\frac{\theta u}{v}\psi'\left(\theta ut/v\right)dt\, h(v,u)dvdu, \end{aligned}$$

for $\theta = \sigma^2/\sigma^{*2} = 1 + r\sigma_a^2/\sigma^2 \geq 1$. Making the transformations $(t/v)u = w$ and $w/t = x$ in order with $(t/v)du = dw$ and $(w/t2)dt = dx$, we observe that the r.h.s. of (5.21) is equal to

$$\begin{aligned} & \int\int\int_1^\infty \left\{v - \frac{1}{\psi(\theta w)}\right\}\frac{\theta w}{t}\psi'\left(\theta w\right)\frac{v}{t}h(v, vw/t)dtdvdw \\ &= \int\int\left\{v - \frac{1}{\psi(\theta w)}\right\}\theta v\psi'\left(\theta w\right)\int_0^w h(v,vx)dxdvdw. \end{aligned} \tag{5.23}$$

Since $\psi'(w) \geq 0$, it is concluded that the r.h.s. of (5.23) is nonnegative if

$$\psi(\theta w) \geq \frac{\int_0^\infty v\int_0^w h(v,vx)dxdv}{\int_0^\infty v^2\int_0^w h(v,vx)dxdv}. \tag{5.24}$$

Since $\theta \geq 1$ and $\psi'(w) \geq 0$, it follows that $\psi(\theta w) \geq \psi(w)$, which, from (5.24), gives the sufficient condition that $\psi(w)$ is greater than or equal to the r.h.s. of (5.24), which, from (5.22), yields the condition (b) of Proposition 5.2, which is established. $\triangle$

The condition of Proposition 5.2 depends on the function g or f through $B_{n+n^*}(g)$. A consequence of Proposition 5.2 presents a class of estimators with robust improvements on $\hat{\sigma}^2_{UB}$, that is, if $\psi(w)$ is nondecreasing and if $\psi(w) \geq \psi_0(w) \sup_g B_{n+n^*}(g)$ with $\lim_{w\to\infty}\psi(w) = n^{-1}$, then $\hat{\sigma}^2(\psi)$ is better than $\hat{\sigma}^2_{UB}$ uniformly within the class of ECD models. It is, nevertheless, difficult to get the value of $\sup_g B_{n+n^*}(g)$, and it may be needed to restrict the class of distributions. One of the restrictions is to impose that

$$e^{w/2}g(w) \text{ is nondecreasing in } w, \tag{5.25}$$

which implies that $B_{n+n^*}(g) \leq (n+n^*)^{-1}$ as shown by the same way as in (5.8). Noting that $\psi_0(w) \leq 1+w$, hence, we get the robust improvements of $\hat{\sigma}^2_{TR}$ upon $\hat{\sigma}^2_{UB}$ for every g satisfying (5.25). The contaminated (or mixture) normal distributions given by (5.9) belong to the class (5.25).

The other choice of distributions with heavier tails is a multivariate t-distribution, where the function $g(z)$ defined by (5.16) is written as

$$g(z) = \frac{\nu^{\nu/2}\Gamma((N-1+\nu)/2)}{\pi^{(N-1)/2}\Gamma(\nu/2)}(\nu+z)^{-(N-1+\nu)/2}, \quad \nu \geq 1, \tag{5.26}$$

for $N-1 = n+n^*$. Since $g(z)$ does not satisfy (5.25), we can directly calculate $B_{n+n^*}(g)$ as

$$B_{n+n^*}(g) = \frac{\int_0^\infty y^{(n+n^*)/2-1}(\nu+y)^{-(n+n^*+\nu)/2}dy}{\int_0^\infty y^{(n+n^*)/2}(\nu+y)^{-(n+n^*+\nu)/2}dy} = \frac{\nu-2}{\nu}\frac{1}{n+n^*} \leq \frac{1}{n+n^*}. \tag{5.27}$$

This demonstrates that $\hat{\sigma}^2_{TR}$ dominates $\hat{\sigma}^2_{UB}$ for every $\nu \geq 1$, that is, this dominance result is robust within the class of multivariate t-distributions.

The robust improvement in estimation of σ^{*2} has a quite different story from the case of estimation of σ^2 although both estimation issues have a similar scenario in the normal distribution. Consider the estimator $\hat{\sigma}^{*2}(\phi) = S^*\phi(S/S^*)$ and evaluate it in terms of the risk

$$R_2(\omega; S^*\phi\left(\frac{S}{S^*}\right)) = E\left[\frac{S^*}{\sigma^{*2}}\phi\left(\frac{S}{S^*}\right) - \log\frac{S^*}{\sigma^{*2}}\phi\left(\frac{S}{S^*}\right) - 1\right].$$

Proposition 5.3. *Assume that*
(a) $\phi(w)$ is nondecreasing and $\phi(0) = a_0^*$,
(b) $\phi(w) \leq \phi_0(w)B_{n+n^*}(g)$ for $B_{n+n^*}(g)$ defined by (5.20), where

$$\phi_0(w) = \frac{\int_w^\infty x^{n/2-1}(1+x)^{-(n+n^*)/2}dx}{\int_w^\infty x^{n/2-1}(1+x)^{-(n+n^*)/2-1}dx}. \tag{5.28}$$

Then $R_2(\omega; S^*\phi(S/S^*)) \leq R_2(\omega; a_0^*S^*)$ uniformly for every ω.

Proof. Since $\phi(0) = a_0^*$, the same arguments as in the proof of Proposition 5.2 give

$$\begin{aligned} & R_2(\omega; a_0^*S^*) - R_2(\omega; S^*\phi\left(\frac{S}{S^*}\right)) \\ &= -E\left[\int_0^1 \frac{d}{dt}\left\{\frac{S^*}{\sigma^{*2}}\phi\left(\frac{S}{S^*}t\right) - \log\frac{S^*}{\sigma^{*2}}\phi\left(\frac{S}{S^*}t\right) - 1\right\}dt\right] \\ &= \int\cdots\int\int_0^1\left\{\frac{1}{\phi(vt/(\theta u))} - u\right\}\frac{v}{\theta u}\phi'\left(\frac{vt}{\theta u}\right)dt\, h(v,u)dvdu \end{aligned} \tag{5.29}$$

for the joint density function $h(v,u)$ given by (5.22). Making the transformations $(t/u)v = w$ and $w(1/t) = y$ in order, we can rewrite (5.29) as

$$\int\int\int_0^1 \left\{\frac{1}{\phi(w/\theta)} - u\right\} \phi'(w/\theta)\frac{uw}{\theta t^2}h(uw/t,u)dtdudw$$
$$= \int\int\int_w^\infty \left\{\frac{1}{\phi(w/\theta)} - u\right\} \phi'(w/\theta)\frac{u}{\theta}h(uy,u)dydudw, \qquad (5.30)$$

so that since $\phi'(w) \geq 0$, the l.h.s. of (5.29) is nonnegative if

$$\phi(w/\theta) \leq \frac{\int\int_w^\infty uh(uy,u)dydu}{\int\int_w^\infty u^2h(uy,u)dydu}. \qquad (5.31)$$

Hence Proposition 5.3 is established by noting that $\phi(w/\theta) \leq \phi(w)$, and that the r.h.s. of (5.31) is equal to $\phi_0(w)B_{n+n^*}(g)$ given in the condition (b). $\triangle$

From Proposition 5.3, we get a class of estimators with robust improvement on $\hat{\sigma}^{*2}_{UB}$, namely, if $\phi(w)$ is nondecreasing and if $\phi(w) \leq \phi_0(w)\inf_g B_{n+n^*}(g)$ with $\phi(0) = n^{*-1}$, then $\hat{\sigma}^{*2}(\phi)$ dominates $\hat{\sigma}^{*2}_{UB}$ uniformly with respect to the ECD model. It is interesting to note here that the required values of $B_{n+n^*}(g)$ defined in (5.20) for the improvement in estimating σ^2 and σ^{*2} are in the opposite directions. For the estimation of $\hat{\sigma}^{*2}$, the value of $\inf_g B_{n+n^*}(g)$ is needed while $\sup_g B_{n+n^*}(g)$ is required for σ^2. As indicated in the above examples, $\sup_g B_{n+n^*}(g)$ is attained when g is the standard normal distribution, namely the distribution with the lightest tail. In contrast with it, $\inf_g B_{n+n^*}(g)$ may be attainable when g has a tail as heavy as possible within a class of distributions under consideration. Two issues of estimations of σ^2 and σ^{*2} thus have different stories in robustness of improvements.

One restriction of distributions corresponding to (5.25) is to impose that for known value $\tau_0 > 1$,

$$e^{w/(2\tau_0)}g(w) \text{ is nonincreasing in } w,$$

which is satisfied by the class (5.9) with the constraint on the dispersion that $\sigma^{*2} \leq \tau_0\sigma^2$. The $\hat{\sigma}^{*2}_{UB}$ is improved on by $\max\{n^{-1}S,\ \tau_0^{-1}(n+n^*)^{-1}(S+S^*)\}$ for every distribution within the restricted class of (5.9).

For the multivariate t-distributions, one needs to impose the restriction on ν as $\nu \geq \nu_0$ for known value $\nu_0 > 2$. Then from (5.27), it follows that $\hat{\sigma}^{*2}_{UB}$ is dominated by $\max\{n^{-1}S, \nu_0^{-1}(\nu_0 - 2)(n+n^*)^{-1}(S+S^*)\}$ uniformly within the class with $\nu \geq \nu_0$.

Apart from the robustness of improvements, we conclude this section with providing decision-theoretic results in the situation where the distribution g is fixed as a known function. Note that unbiased (or best scale equivariant) estimators of σ^2 and σ^{*2} are given by $\hat{\sigma}^2_{UB}(g) = n^{-1}(n+n^*)B_{n+n^*}(g)$ and $\hat{\sigma}^{*2}_{UB}(g) = n^{*-1}(n+n^*)B_{n+n^*}(g)$. Proposition 5.2 shows that $\hat{\sigma}^2_{UB}(g)$ is dominated by smooth estimator $\hat{\sigma}^2_{GB}(g) = S\psi_0(S^*/S)B_{n+n^*}(g)$ and truncated estimator $\hat{\sigma}^2_{EB}(g) = \min\{n^{-1}S,\ (n+n^*)^{-1}(S+S^*)\}(n+n^*)B_{n+n^*}(g)$. It is here interesting to note that $\hat{\sigma}^2_{GB}(g)$ is interpreted as a generalized Bayes estimator against the prior distribution $\eta^{-1}\xi^{-1}d\eta d\xi$, $0 < \xi < 1$, for $\eta = 1/\sigma^2$ and $\xi = \sigma^2/\sigma^{*2}$. Also $\hat{\sigma}^2_{EB}(g)$ is derived as an empirical Bayes estimator against the prior distribution $\eta^{-1}d\eta$ with unknown parameter ξ, $0 < \xi < 1$. In fact, the Bayes estimator of σ^2 is of the form $\hat{\sigma}^2_B(\xi) = (S+\xi S^*)B_{n+n^*}(g)$, and the marginal density of (S, S^*) is given by $(const)\xi^{n^*/2}(S+\xi S^*)^{-(n+n^*)/2}\int \eta^{(n+n^*)/2-1}g(\eta)d\eta$. The maximum likelihood estimator of ξ is $\min\{n^*S/(nS^*),\ 1\}$, which is substituted in $\hat{\sigma}^2_B(\xi)$ and we get the empirical Bayes estimator $\hat{\sigma}^2_{EB}(g)$. For estimation of σ^{*2}, the same arguments and Proposition 5.3 provide the generalized Bayes estimator $\hat{\sigma}^{*2}_{GB}(g) = S^*\phi_0(S/S^*)B_{n+n^*}(g)$ and the empirical Bayes estimator $\hat{\sigma}^{*2}_{EB}(g) = \min\{n^{*-1}S,\ (n+n^*)^{-1}(S+S^*)\}(n+n^*)B_{n+n^*}(g)$, both improving on $\hat{\sigma}^{*2}_{UB}(g)$. For instance, $B_{n+n*}(g) = (n+n^*)^{-1}$ for the normal distribution while $B_{n+n^*}(g) = \nu^{-1}(\nu-2)(n+n^*)^{-1}$ for the multivariate t-distribution.

Acknowledgements
The research of the first author was supported in part by a grant from the Shirakawa Institute of Animal Genetics and the Japan Livestock Technology Association, by the Ministry of Education, Japan, Grant No. 08780216, No. 09780214 and by a grant from the Research Institute for the Japanese Economy, University of Tokyo. The research of the second author was supported by Natural Science and Engineering Research Council of Canada.

REFERENCES

Bilodeau, M. and Kariya, T. (1989). Minimax estimators in the normal MANOVA model. *J. Multivariate Anal.*, **28**, 260-270.

Brandwein, A.C. and Strawderman, W.E. (1990). Stein estimation: The spherically symmetric case. *Statist. Science*, **5**, 356-369.

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.

Cellier, D. and Fourdrinier, D. (1995a). Shrinkage estimators under spherical symmetry for the general linear model. *J. Multivariate Anal.*, **52**, 338-351.

Cellier, D. and Fourdrinier, D. (1995b). Shrinkage positive rule estimators for spherically symmetric distributions. *J. Multivariate Anal.*, **53**, 194-209.

Dey, D.K. and Srinivasan, C. (1985). Estimation of covariance matrix under Stein's loss. *Ann. Statist.*, **13**, 1581-1591.

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

Gleser, L.J. (1986). Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.*, **14**, 1625-1633.

Haff, L.R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, **9**, 531-542.

Honda, T. (1991). Minimax estimators in the MANOVA model for arbitrary quadratic loss and unknown covariance matrix. *J. Multivariate Anal.*, **36**, 113-120.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, **1**, 361-379. University of California Press, Berkeley.

Kariya, T., Konno, Y. and Strawderman, W.E. (1996). Double shrinkage estimators in the GMANOVA model. *J. Multivariate Anal.*, **56**, 245-258.

Konno, Y. (1990). On estimating the matrix of mean. Unpublished manuscript.

Konno, Y. (1991). On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivariate Anal.*, **36**, 44-55.

Konno, Y., Kariya, T. and Strawderman, W.E. (1994). *Construction of improved estimators in the regression coefficient matrix for the GMANOVA model.* Discussion Paper Series A, Inst. Econ. Res., Hitotsubashi University.

Kubokawa, T.(1994). A unified approach to improving equivariant estimators. *Ann. Statist.* **22**, 290-299.

Kubokawa, T. (1995). Estimation of variance components in mixed linear models. *J. Multivariate Anal.*, **53**, 210-236.

Kubokawa, T.(1997). *The Stein phenomenon in simultaneous estimation: A review.* Discussion Paper Series 97-F-14, Faculty of Economics, University of Tokyo.

Kubokawa, T. and Srivastava, M.S. (1996). Double shrinkage estimators of ratio of variances. In *Multidimensional Statistical Analysis and Theory of Random Matrices* (Proceedings of the Sixth Eugene Lukacs Symposium, Bowling Green, OH, USA, 29-30 March 1996, Editors: A.K. Gupta and V.L. Girko), 139-154, VSP, Netherlands.

Robert, C.P. (1994). *The Bayesian Choice : A Decision-Theoretic Motivation.* Springer-Verlag, 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.

Srivastava, M.S. and Khatri, C.G. (1979). *An Introduction to Multivariate Statistics.* North-Holland, New York.

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. University of California University, Berkeley.

Stein, C.(1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.*, **16**, 155-160.

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

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

Stein, C. (1977b). Lectures on multivariate estimation theory. In *Investigation on Statistical Estimation Theory* I, 4-65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad.(in Russian)

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

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)

Tan, M. (1991). Improved estimators for the GMANOVA problem with application to Monte Carlo simulation. *J. Multivariate Anal.*, **38**, 262-274.

Takeuchi, K.(1991). Personal Communications.