Ann. Inst. Statist. Math.
Vol. 56, No. 4, 757–770 (2004)

NEW ESTIMATORS OF DISCRIMINANT COEFFICIENTS AS THE GRADIENT OF LOG-ODDS

YO SHEENA* AND ARJUN K. GUPTA

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, U.S.A.

(Received June 5, 2003; revised February 23, 2004)

Abstract. We consider the problem of estimating the discriminant coefficients, $\boldsymbol{\eta} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})$ based on two independent normal samples from $N_p(\boldsymbol{\theta}^{(1)}, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\theta}^{(2)}, \boldsymbol{\Sigma})$. We are concerned with the estimation of $\boldsymbol{\eta}$ as the gradient of log-odds between two extreme situations. A decision theoretic approach is taken with the quadratic loss function. We derive the unbiased estimator of the essential part of the risk which is applicable for general estimators. We propose two types of new estimators and prove their dominance over the traditional estimator using this unbiased estimator.

Key words and phrases: Unbiased estimator of risk, linear discriminant function, posterior log-odds.

1. Introduction

Let X_{1i} $(i = 1, \dots, n_1)$ and X_{2i} $(i = 1, \dots, n_2)$ be two samples of training data which were drawn from populations $P_1, N_p(\boldsymbol{\theta}^{(1)}, \boldsymbol{\Sigma})$ and $P_2, N_p(\boldsymbol{\theta}^{(2)}, \boldsymbol{\Sigma})$ respectively. Then $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})$ is called the discriminant coefficient. Fisher's linear discriminant function of $\boldsymbol{x}$, a p-dimensional new data set, is given by $\boldsymbol{x}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})$.

As Haff (1986) observed, the discriminant coefficient is the gradient of posterior log-odds, which itself is a meaningful subject of estimation. We briefly mention here how the discriminant coefficient can be practically used. In the "discriminant analysis" in its original or strict meaning, we are given two mutually exclusive groups and the interest is in the classification of a new data, where some examples of possible questions are: "Which species does this iris belong to?"; "Is this tumor malignant or benign?"; "Who wrote this document, James Madison or Alexander Hamilton?". On the other hand, we often encounter the situations where a new data can hold a position anywhere between two extreme groups. Consider for example two extreme groups such as: financially excellent companies and those which went bankrupt; students who graduate cum laude and those who drop out. In these situations the main concern would be the relative closeness of a new data (company/student) to the two extreme groups. Temporarily we call it a "placement" problem in order to distinguish it from the "discrimination" problem in the strict sense. Since the statistical tools are quite similar for the two problems, they have

*On leave from Shinshu University, Japan. Present corresponding address: Department of Economics, Shinshu University, 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan, e-mail: sheena@econ.shinshu-u.ac.jp

been treated uniformly under the same name such as classification or discrimination. Actually in many texts on multivariate analysis, examples of the two problems are taken as an application of the discriminant analysis. See for example Johnson and Wichern (1998).

However the interpretation or usage of a discriminant function is naturally different in the two problems. In the "discrimination" problem, it is used as a tool for the classification of a new data, while in the "placement" problem it is used for the placement of a new data between two extreme groups. We will illustrate this point with Altman's Z-score. Altman (1968) applied Fisher's linear discriminant function to the financial data from two groups of companies: companies which had gone bankrupt and those which still existed in a particular year. The updated version of Z-score for a private company (Altman (1993)) is given by

$$Z = 6.72x_1 + 3.26x_2 + 6.56x_3 + 1.05x_4,$$

where x_1 = (earnings before taxes + interest)/total assets, x_2 = retained earnings/total assets, x_3 = working capital/total assets, x_4 = market value of equity/total liabilities. (See also Grice and Ingram (2001) for Altman's Z.) The coefficients x_i, $i = 1, \ldots, 4$ are discriminant coefficients calculated from the sample of the two groups of companies. Originally Z-score was proposed to "discriminate" the two groups, bankrupt and non-bankrupt firms. However, it has been practically used by many financial analysts to assess bankruptcy potential of firms. (You can find hundreds of Web Pages to provide consultant service using this score.) In other words, Z-score is the indicator of a company's financial healthiness. The two groups of companies, "excellent" companies and those filed a bankruptcy petition are two extreme groups (training data) and a company of concern is evaluated by its Z-score, the relative closeness to those points so that the company is "placed" somewhere between the two groups. The relative closeness can be given by the likelihood ratio between the distributions of each group. If we assume these distributions are $N_p(\boldsymbol{\theta}^{(1)}, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\theta}^{(2)}, \boldsymbol{\Sigma})$, the log-likelihood ratio equals to $\boldsymbol{x}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}) + a_0$ with some constant a_0. Then the estimation of the discriminant coefficient and a_0 is needed. Especially the discriminant coefficient, the gradient of the log-likelihood ratio is important, since the variation over a certain time span (like annual or quarterly movement) is determined by the gradient.

In this paper we consider the estimation of the discriminant coefficient in a "placement" problem. It is most reasonable to evaluate an estimator of the discriminant coefficient by the probability of misclassification, if the case is the "discrimination" problem. However for "placement" problem, there is no such reasonable criteria. Therefore we use the risk with respect to a quadratic loss function for the evaluation of estimators. Our aim is to present new estimators which dominate a traditional estimator with respect to the risk.

By the usual canonical reduction of the problem, we can assume that $\boldsymbol{Y} = (y_1, \ldots, y_p)^t$ and $\boldsymbol{S} = (s_{ij})$ are independently distributed as

$$\boldsymbol{Y} \sim N_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}), \qquad \boldsymbol{S} \sim \boldsymbol{W}_p(k, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\theta} = m(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})$ with $m = (n_1 n_2/(n_1 + n_2))^{1/2}$ and $k = n_1 + n_2 - 2$. We assume $k > p + 1$ unless otherwise stated. We estimate $\boldsymbol{\eta} = m\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)}) = \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}$ instead of $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^{(1)} - \boldsymbol{\theta}^{(2)})$ for simplicity, based on $\boldsymbol{Y}$ and $\boldsymbol{S}$. We evaluate an estimator

$\widehat{\boldsymbol{\eta}} = (\widehat{\eta}_1, \ldots, \widehat{\eta}_p)^t$ in view of its risk using the loss function

$$\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|^2 = (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = \sum_{i=1}^{p}(\widehat{\eta}_i - \eta_i)^2. \tag{1.1}$$

Hereafter we use the notation, $R(\widehat{\boldsymbol{\eta}}; \boldsymbol{\theta}, \boldsymbol{\Sigma})$ for the risk of $\widehat{\boldsymbol{\eta}}$, i.e., $R(\widehat{\boldsymbol{\eta}}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) = E[\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|^2]$.

Haff (1986) originally considered the estimation of $\boldsymbol{\eta}$ under the loss function (1.1). Later Dey and Srinivasan (1991) considered the estimation of $\boldsymbol{\eta}$ in the same framework. Sarkar and Krishnamoorthy (1991) and Rukhin (1992) also deal with the same problem from a decision theoretic point of view, but they used another type of loss function,

$$(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})^t \boldsymbol{\Sigma} (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}). \tag{1.2}$$

Sarkar and Krishnamoorthy (1991) proved the dominance result of Stein-type estimators over a traditional estimator and Rukhin (1992) derived generalized Bayes estimators.

A traditional estimator, which is commonly used in practice, is the unbiased estimator,

$$\widehat{\boldsymbol{\eta}}^{(u)} = (k - p - 1)\boldsymbol{S}^{-1}\boldsymbol{Y}. \tag{1.3}$$

Haff (1986) (see Lemma 4.4) and Dey and Srinivasan (1991) (see Solution 2) proved that a simple shrinkage estimator,

$$\widehat{\boldsymbol{\eta}}^{(H_1)} = a\boldsymbol{S}^{-1}\boldsymbol{Y}, \qquad k - p - 5 < a < k - p - 1 \tag{1.4}$$

dominates $\widehat{\boldsymbol{\eta}}^{(u)}$ w.r.t. the loss function, (1.1). Haff proposed another estimator of the form

$$\widehat{\boldsymbol{\eta}}^{(H_2)} = (k - p - 1)(\boldsymbol{S} + ut(u)\boldsymbol{I}_p)^{-1}\boldsymbol{Y}, \tag{1.5}$$

where $u = 1/\operatorname{tr}\boldsymbol{S}^{-1}$ and $t(u)$ is a positive valued absolutely continuous function of u. He proved that $\widehat{\boldsymbol{\eta}}^{(H_2)}$ dominates $\widehat{\boldsymbol{\eta}}^{(u)}$ w.r.t. the loss function (1.1) if $(k - p - 5)t^2(u) - 4t(u) + 4ut'(u) \le 0$, $\forall u > 0$ and some conditions are met for the application of Stokes' theorem.

Dey and Srinivasan (1991) (see Solution 3) proposed an estimator of the form

$$\widehat{\boldsymbol{\eta}}^{(DS)} = a\boldsymbol{S}^{-1}\boldsymbol{Y} + b(\operatorname{tr}\boldsymbol{S})^{-1}\boldsymbol{Y} \tag{1.6}$$

and proved that $\widehat{\boldsymbol{\eta}}^{(DS)}$ dominates $\widehat{\boldsymbol{\eta}}^{(u)}$ w.r.t. the loss function (1.1) if $k - p - 5 < a < k - p - 3$, $0 < b \le 2(k - p - 1 - a) - 4$ and some conditions are met for the application of Stokes' theorem.

In the next section we describe the unbiased estimator of (a part of) the risk, $R(\widehat{\boldsymbol{\eta}}; \boldsymbol{\theta}, \boldsymbol{\Sigma})$. In Sections 3 and 4, we present new estimators and prove the dominance result of these estimators over $\widehat{\boldsymbol{\eta}}^{(u)}$ using the unbiased estimator of the risk.

2. Unbiased estimator of risk

First we derive the unbiased estimator of an essential part of $R(\widehat{\boldsymbol{\eta}}; \boldsymbol{\theta}, \boldsymbol{\Sigma})$ for a general estimator $\widehat{\boldsymbol{\eta}}(\boldsymbol{Y}, \boldsymbol{S})$. Haff (1986) derived the same kind of risk formula for the estimator of the form $\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{S})\boldsymbol{Y}$. We present a new formula which is applicable for a general class of estimators.

For the purpose of risk comparison between two estimators, the last term, $\boldsymbol{\eta}^t\boldsymbol{\eta}$ of the expanded loss function,

$$\|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|^2 = \widehat{\boldsymbol{\eta}}^t\widehat{\boldsymbol{\eta}} - 2\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}} + \boldsymbol{\eta}^t\boldsymbol{\eta}$$

is irrelevant. Therefore we focus on the first two terms. Let

$$R^*(\widehat{\boldsymbol{\eta}}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) = E[\widehat{\boldsymbol{\eta}}^t\widehat{\boldsymbol{\eta}}] - 2E[\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}}].$$

Before stating the formula on $E[\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}}]$ in Lemma 2.1, we enumerate all the conditions required for the proof. Most of these conditions are concerned with the evaluation of the integrals on boundaries that naturally arise in the application of Stokes' theorem. Haff (1986) and Dey and Srinivasan (1991) does not mention these conditions explicitly, but we think it is better for further usage of Lemma 2.1 to describe them explicitly. Let

$$\begin{aligned} d\boldsymbol{Y}_{\hat{i}} &= dy_1 \cdots dy_{i-1}dy_{i+1} \cdots dy_p, \\ \partial S_1(c,d) &= \{\boldsymbol{S} > 0 \mid \operatorname{tr} \boldsymbol{S} = c \text{ and } \det \boldsymbol{S} \geq d\}, \\ \partial S_2(c,d) &= \{\boldsymbol{S} > 0 \mid \operatorname{tr} \boldsymbol{S} \leq c \text{ and } \det \boldsymbol{S} = d\}. \end{aligned}$$

Then the conditions are as follows:

(1) $\widehat{\eta}_i$ $(i = 1, \ldots, p)$ is a C^1 function of $\boldsymbol{Y}$ and $\boldsymbol{S}$;

(2) $\lim_{(c,d)\to(\infty,0)} \int_{\partial S_1(c,d)} \exp(-\frac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}\boldsymbol{S})|\boldsymbol{S}|^{(k-p-1)/2}\widehat{\eta}_j(\boldsymbol{Y}, \boldsymbol{S}) \bigwedge_{(l,m)\neq(i,j)} ds_{lm} = 0$, $\forall \boldsymbol{Y} \in \Re^p$, $1 \leq \forall i \leq \forall j \leq p$;

(3) $\lim_{(c,d)\to(\infty,0)} \int_{\partial S_2(c,d)} \exp(-\frac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}\boldsymbol{S})|\boldsymbol{S}|^{(k-p-1)/2}\widehat{\eta}_j(\boldsymbol{Y}, \boldsymbol{S}) \bigwedge_{(l,m)\neq(i,j)} ds_{lm} = 0$, $\forall \boldsymbol{Y} \in \Re^p$, $1 \leq \forall i \leq \forall j \leq p$;

(4) $\lim_{c\to\infty} \int_{\Re^{p-1}} [\exp(-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{\theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\theta}))\widehat{\eta}_i(\boldsymbol{Y}, \boldsymbol{S})]_{y_i=-c}^{y_i=c} d\boldsymbol{Y}_{\hat{i}} = 0$, $\forall \boldsymbol{S} > 0$, $1 \leq \forall i \leq p$;

(5) In the following proof of Lemma 2.1, the left side integrals in (2.2) and the subsequent limit operation are exchangeable. The exchangeability also holds between the integral on the left side of (2.5) and the subsequent limit operation.

LEMMA 2.1.

$$E[\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}}] = E\left[2\boldsymbol{Y}^t\boldsymbol{D}\widehat{\boldsymbol{\eta}} + (k-p-1)\boldsymbol{Y}^t\boldsymbol{S}^{-1}\widehat{\boldsymbol{\eta}} - \sum_{i=1}^p \frac{\partial\widehat{\eta}_i}{\partial y_i}\right],$$

where

$$\boldsymbol{D} = (d_{ij}), \qquad d_{ij} = \frac{1}{2}(1 + \delta_{ij})\frac{\partial}{\partial s_{ij}}$$

and

$$\boldsymbol{D}\widehat{\boldsymbol{\eta}} = \left(\frac{1}{2}\sum_{j=1}^p (1 + \delta_{1j})\frac{\partial\widehat{\eta}_j}{\partial s_{1j}}, \ldots, \frac{1}{2}\sum_{j=1}^p (1 + \delta_{pj})\frac{\partial\widehat{\eta}_j}{\partial s_{pj}}\right)^t.$$

PROOF.

$$E[\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}}] = E_S[E_Y[\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}} \mid \boldsymbol{S}]]$$

and the inner expectation equals to

$$K_0 \int_{\Re^p} \exp\left(-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\theta})\right)\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}}d\boldsymbol{Y} = K_0 \int_{\Re^p} \exp\left(-\frac{1}{2}\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}+\boldsymbol{\eta}^t\boldsymbol{Y}-\frac{1}{2}\boldsymbol{\theta}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\right)\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}}d\boldsymbol{Y}, \tag{2.1}$$

where K_0 is some constant. Since

$$\frac{\partial}{\partial y_i}\exp\left(-\frac{1}{2}\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}+\boldsymbol{\eta}^t\boldsymbol{Y}-\frac{1}{2}\boldsymbol{\theta}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\right) = \left(-\sum_{j=1}^p \sigma^{ij}y_j+\eta_i\right)\exp\left(-\frac{1}{2}\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}+\boldsymbol{\eta}^t\boldsymbol{Y}-\frac{1}{2}\boldsymbol{\theta}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\right) \quad (\sigma^{ij}=(\boldsymbol{\Sigma}^{-1})_{ij}),$$

we have, by integral by parts,

$$\begin{aligned}&\int_{\Re^{p-1}}\int_{-c}^{c}\exp\left(-\frac{1}{2}\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}+\boldsymbol{\eta}^t\boldsymbol{Y}-\frac{1}{2}\boldsymbol{\theta}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\right)\\ &\qquad\times\left(-\sum_{j=1}^p\sigma^{ij}y_j+\eta_i\right)\widehat{\eta}_i dy_i d\boldsymbol{Y}_{\hat{i}}\\ &+\int_{\Re^{p-1}}\int_{-c}^{c}\exp\left(-\frac{1}{2}\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}+\boldsymbol{\eta}^t\boldsymbol{Y}-\frac{1}{2}\boldsymbol{\theta}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}\right)\frac{\partial\widehat{\eta}_i}{\partial y_i}dy_i d\boldsymbol{Y}_{\hat{i}}\\ &=\int_{\Re^{p-1}}\left[\exp\left(-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\theta})\right)\widehat{\eta}_i\right]_{y_i=-c}^{y_i=c}d\boldsymbol{Y}_{\hat{i}}.\end{aligned} \tag{2.2}$$

Consider the limit values of both sides of (2.2) as c goes to infinity. Then conditions (4) and (5) gives

$$E_Y[\eta_i\widehat{\eta}_i \mid \boldsymbol{S}] = E_Y\left[\widehat{\eta}_i\sum_{j=1}^p\sigma^{ij}y_j-\frac{\partial\widehat{\eta}_i}{\partial y_i}\,\middle|\,\boldsymbol{S}\right].$$

Hence

$$E_Y[\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}}\mid\boldsymbol{S}] = E_Y\left[\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\eta}}-\sum_{i=1}^p\frac{\partial\widehat{\eta}_i}{\partial y_i}\,\middle|\,\boldsymbol{S}\right]$$

and

$$E[\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}}] = E\left[\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\eta}}-\sum_{i=1}^p\frac{\partial\widehat{\eta}_i}{\partial y_i}\right]. \tag{2.3}$$

Now we evaluate $E_S[\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\eta}}\mid\boldsymbol{Y}]$ using Stokes' theorem. The same result would be gained by the straightforward application of Haff's Wishart identity (See Haff (1982)), but in order to make clear the conditions to be satisfied by the estimator, we describe how Stokes' theorem is applied to the evaluation of $E_S[\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\eta}}\mid\boldsymbol{Y}]$.

With some constant K_1, we have

$$(2.4)\qquad \begin{aligned} &E_S[\sigma^{ij}\widehat{\eta}_j \mid \boldsymbol{Y}] \\ &= K_1 \int_{S>0} \exp\left(-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)|\boldsymbol{S}|^{(k-p-1)/2}\widehat{\eta}_j\sigma^{ij}d\boldsymbol{S}, \qquad 1 \le i \le j \le p. \end{aligned}$$

Note that

$$\begin{aligned} &-(1+\delta_{ij})\frac{\partial}{\partial s_{ij}}\exp\left(-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) \\ &= -(1+\delta_{ij})\frac{\partial}{\partial s_{ij}}\exp\left(-\sum_{i\le j}^{p}\frac{\sigma^{ij}}{1+\delta_{ij}}s_{ij}\right) \qquad (\delta_{ij}\text{: Kronecker's delta}) \\ &= \exp\left(-\sum_{i\le j}^{p}\frac{\sigma^{ij}}{1+\delta_{ij}}s_{ij}\right)\sigma^{ij}. \end{aligned}$$

By Stokes' theorem, we have

$$(2.5)\qquad \begin{aligned} &\int_{\mathcal{S}(c,d)}\frac{\partial}{\partial s_{ij}}\left(\exp\left(-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)|\boldsymbol{S}|^{(k-p-1)/2}\widehat{\eta}_j\right)\bigwedge_{l\le m} ds_{lm} \\ &= \int_{\partial\mathcal{S}_1(c,d)}\exp\left(-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)|\boldsymbol{S}|^{(k-p-1)/2}\widehat{\eta}_j \bigwedge_{(l,m)\ne(i,j)} ds_{lm} \\ &\quad + \int_{\partial\mathcal{S}_2(c,d)}\exp\left(-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)|\boldsymbol{S}|^{(k-p-1)/2}\widehat{\eta}_j \bigwedge_{(l,m)\ne(i,j)} ds_{lm}, \end{aligned}$$

where

$$\mathcal{S}(c,d) = \{\boldsymbol{S} > 0 \mid \operatorname{tr}\boldsymbol{S} \le c \text{ and } \det\boldsymbol{S} \ge d\}$$

and $\partial\mathcal{S}_i(c,d)$, $i = 1,2$ was already defined when we stated the conditions of the lemma. Take the limit on both sides of (2.5) as $(c,d) \to (\infty, 0)$. The conditions (2), (3) and (5) and the fact

$$\frac{\partial|\boldsymbol{S}|}{\partial s_{ij}} = \frac{2}{1+\delta_{ij}}s^{ij}|\boldsymbol{S}| \qquad (s^{ij} = (\boldsymbol{S}^{-1})_{ij})$$

lead us to

$$(2.6)\qquad \begin{aligned} &\int_{S>0}\exp\left(-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)|\boldsymbol{S}|^{(k-p-1)/2}\widehat{\eta}_j\sigma^{ij}d\boldsymbol{S} \\ &= (k-p-1)\int_{S>0}\exp\left(-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)|\boldsymbol{S}|^{(k-p-1)/2}s^{ij}\widehat{\eta}_j d\boldsymbol{S} \\ &\quad + (1+\delta_{ij})\int_{S>0}\exp\left(-\frac{1}{2}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)|\boldsymbol{S}|^{(k-p-1)/2}\frac{\partial\widehat{\eta}_j}{\partial s_{ij}}d\boldsymbol{S}. \end{aligned}$$

From (2.4) and (2.6), we have

$$E_S[\sigma^{ij}\widehat{\eta}_j \mid \boldsymbol{Y}] = E_S\left[(k-p-1)s^{ij}\widehat{\eta}_j + (1+\delta_{ij})\frac{\partial\widehat{\eta}_j}{\partial s_{ij}} \;\middle|\; \boldsymbol{Y}\right], \qquad 1 \le \forall i, \forall j \le p,$$

and

$$\begin{aligned} &E_S[\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\eta}} \mid \boldsymbol{Y}] \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} E_S[y_i\sigma^{ij}\widehat{\eta}_j \mid \boldsymbol{Y}] \\ &= E_S[(k-p-1)\boldsymbol{Y}^t\boldsymbol{S}^{-1}\widehat{\boldsymbol{\eta}} + 2\boldsymbol{Y}^t\boldsymbol{D}\widehat{\boldsymbol{\eta}} \mid \boldsymbol{Y}]. \end{aligned}$$

Therefore we have

$$E[\boldsymbol{Y}^t\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\eta}}] = E[(k-p-1)\boldsymbol{Y}^t\boldsymbol{S}^{-1}\widehat{\boldsymbol{\eta}} + 2\boldsymbol{Y}^t\boldsymbol{D}\widehat{\boldsymbol{\eta}}] \tag{2.7}$$

and from (2.3) and (2.7), we get

$$E[\boldsymbol{\eta}^t\widehat{\boldsymbol{\eta}}] = E\left[2\boldsymbol{Y}^t\boldsymbol{D}\widehat{\boldsymbol{\eta}} + (k-p-1)\boldsymbol{Y}^t\boldsymbol{S}^{-1}\widehat{\boldsymbol{\eta}} - \sum_{i=1}^{p}\frac{\partial\widehat{\eta}_i}{\partial y_i}\right]. \qquad \square$$

Further from Lemma 2.1, we have

$$R^*(\widehat{\boldsymbol{\eta}};\boldsymbol{\theta},\boldsymbol{\Sigma}) = E\left[\widehat{\boldsymbol{\eta}}^t\widehat{\boldsymbol{\eta}} - 4\boldsymbol{Y}^t\boldsymbol{D}\widehat{\boldsymbol{\eta}} - 2(k-p-1)\boldsymbol{Y}^t\boldsymbol{S}^{-1}\widehat{\boldsymbol{\eta}} + 2\sum_{i=1}^{p}\frac{\partial\widehat{\eta}_i}{\partial y_i}\right]. \tag{2.8}$$

3. New estimator I

In this section we consider a simple shrinkage estimator that dominates the unbiased estimator $\widehat{\boldsymbol{\eta}}^{(u)}$. The coefficient a in $\widehat{\boldsymbol{\eta}}^{(H_1)}$ is constant. We propose a new estimator which shrinks $\widehat{\boldsymbol{\eta}}^{(u)}$ according to $\boldsymbol{Y}$ and $\boldsymbol{S}$, i.e.,

$$\widehat{\boldsymbol{\eta}}^{(1)} = (\widehat{\eta}_1^{(1)},\dots,\widehat{\eta}_p^{(1)})^t = a(\boldsymbol{Y},\boldsymbol{S})\boldsymbol{S}^{-1}\boldsymbol{Y}. \tag{3.1}$$

We have the following result for this type of estimator.

LEMMA 3.1. *Suppose* $\widehat{\boldsymbol{\eta}}^{(1)}$ *satisfies all the conditions from* (1) *to* (5). *Then we have*

$$\begin{aligned} R^*(\widehat{\boldsymbol{\eta}}^{(1)};\boldsymbol{\theta},\boldsymbol{\Sigma}) = E[&(a^2-2(k-p-2)a)\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y} - 2\boldsymbol{Y}^t\boldsymbol{A}^*\boldsymbol{S}^{-1}\boldsymbol{Y} \\ &+ 2a(\operatorname{tr}\boldsymbol{S}^{-1})\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y} + 2(\operatorname{grad} a)^t\boldsymbol{S}^{-1}\boldsymbol{Y} + 2a\operatorname{tr}\boldsymbol{S}^{-1}], \end{aligned}$$

where

$$(\boldsymbol{A}^*)_{ij} = (1+\delta_{ij})\frac{\partial a}{\partial s_{ij}} \quad \textit{and} \quad \operatorname{grad} a = \left(\frac{\partial a}{\partial y_1},\dots,\frac{\partial a}{\partial y_p}\right)^t.$$

PROOF. From (2.8), we have

$$\begin{aligned} &R^*(\widehat{\boldsymbol{\eta}}^{(1)};\boldsymbol{\theta},\boldsymbol{\Sigma}) \\ &= E\left[(\widehat{\boldsymbol{\eta}}^{(1)})^t\widehat{\boldsymbol{\eta}}^{(1)} - 4\boldsymbol{Y}^t\boldsymbol{D}\widehat{\boldsymbol{\eta}}^{(1)} - 2(k-p-1)\boldsymbol{Y}^t\boldsymbol{S}^{-1}\widehat{\boldsymbol{\eta}}^{(1)} + 2\sum_{i=1}^{p}\frac{\partial\widehat{\eta}_i^{(1)}}{\partial y_i}\right]. \end{aligned} \tag{3.2}$$

Since

$$\text{(3.3)} \qquad \boldsymbol{D}\boldsymbol{S}^{-1} = -\frac{1}{2}((\operatorname{tr}\boldsymbol{S}^{-1})\boldsymbol{S}^{-1} + \boldsymbol{S}^{-2}) \qquad \text{(see Lemma 6 (i) of Haff (1982))},$$

we have

$$\text{(3.4)} \qquad \begin{aligned} \boldsymbol{D}\widehat{\boldsymbol{\eta}}^{(1)} &= (\boldsymbol{D}a\boldsymbol{I}_p)\boldsymbol{S}^{-1}\boldsymbol{Y} + a(\boldsymbol{D}\boldsymbol{S}^{-1})\boldsymbol{Y} \\ &= \frac{1}{2}\boldsymbol{A}^*\boldsymbol{S}^{-1}\boldsymbol{Y} - \frac{a}{2}((\operatorname{tr}\boldsymbol{S}^{-1})\boldsymbol{S}^{-1} + \boldsymbol{S}^{-2})\boldsymbol{Y}, \end{aligned}$$

while

$$\text{(3.5)} \qquad \begin{aligned} \sum_{i=1}^{p} \frac{\partial \widehat{\eta}_i^{(1)}}{\partial y_i} &= \sum_{i=1}^{p} \frac{\partial}{\partial y_i}\left(a\sum_{m=1}^{p} s^{im} y_m\right) \\ &= \sum_{i=1}^{p} \frac{\partial a}{\partial y_i}\sum_{m=1}^{p} s^{im} y_m + a\sum_{i=1}^{p}\sum_{m=1}^{p} s^{im}\frac{\partial y_m}{\partial y_i} \\ &= (\operatorname{grad} a)^t \boldsymbol{S}^{-1}\boldsymbol{Y} + a\operatorname{tr}\boldsymbol{S}^{-1}. \end{aligned}$$

If we substitute (3.4) and (3.5) in (3.2), we have the result. □

Now we consider more specific estimators. Suppose $a(\boldsymbol{Y}, \boldsymbol{S})$ of $\widehat{\boldsymbol{\eta}}^{(1)}$ depends on $\boldsymbol{Y}$ and $\boldsymbol{S}$ only through $z = \boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y}$, i.e.,

$$\text{(3.6)} \qquad \widehat{\boldsymbol{\eta}}^{(1)} = a(z)\boldsymbol{S}^{-1}\boldsymbol{Y}.$$

We prove the dominance of $\widehat{\boldsymbol{\eta}}^{(1)}$ given by (3.6) over $\widehat{\boldsymbol{\eta}}^{(u)}$. The following theorem is the extension of the result gained by Haff (1986) and Dey and Srinivasan (1991).

THEOREM 3.1. *Let $k > p + 5$. Suppose that $a(z)$ and its derivative $a'(z)$ are bounded continuous functions of z and that $k - p - 1 > a(z) \geq k - p - 5$ and $a'(z) \leq 0$. Then $\widehat{\boldsymbol{\eta}}^{(1)}$ in* (3.6) *dominates $\widehat{\boldsymbol{\eta}}^{(u)}$ w.r.t. the loss function* (1.1).

PROOF. It is not difficult but somewhat tedious to show $\widehat{\boldsymbol{\eta}}^{(1)}$ satisfies all the conditions from (1) to (5) using the fact $a(z)$ and $a'(z)$ are bounded continuous functions. Hence we omit it and start argument with the application of Lemma 3.1 to the estimator $\widehat{\boldsymbol{\eta}}^{(1)}$ in (3.6).

Using

$$\frac{\partial s^{lm}}{\partial s_{ij}} = -\frac{1}{1+\delta_{ij}}(s^{li}s^{jm} + s^{lj}s^{im}),$$

we have

$$\text{(3.7)} \qquad \begin{aligned} (\boldsymbol{A}^*)_{ij} &= (1+\delta_{ij})a'(z)\frac{\partial z}{\partial s_{ij}} \\ &= -a'(z)\boldsymbol{Y}^t\boldsymbol{S}^{(ij)}\boldsymbol{Y}, \end{aligned}$$

where $\boldsymbol{S}^{(ij)}$ $(1 \leq i, j \leq p)$ is the $p \times p$ symmetric matrix defined by

$$(\boldsymbol{S}^{(ij)})_{lm} = -(1+\delta_{ij})\frac{\partial s^{lm}}{\partial s_{ij}} = s^{li}s^{jm} + s^{lj}s^{im}.$$

Starting from (3.7), by straightforward calculation, we have

$$\boldsymbol{Y}^t\boldsymbol{A}^*\boldsymbol{S}^{-1}\boldsymbol{Y} = -2a'(z)(\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y})(\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y}). \tag{3.8}$$

Furthermore the term $(\operatorname{grad} a(z))^t\boldsymbol{S}^{-1}\boldsymbol{Y}$ turns out to be

$$\begin{aligned}(\operatorname{grad} a(z))^t\boldsymbol{S}^{-1}\boldsymbol{Y} &= a'(z)\left(\frac{\partial z}{\partial y_1}, \dots, \frac{\partial z}{\partial y_p}\right)\boldsymbol{S}^{-1}\boldsymbol{Y} \\ &= 2a'(z)\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y}.\end{aligned} \tag{3.9}$$

If we substitute (3.8) and (3.9) in the equation of Lemma 3.1, we have

$$\begin{aligned}R^*(\widehat{\boldsymbol{\eta}}^{(1)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) = E[&(a^2 - 2(k-p-2)a)\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y} + 2a(\operatorname{tr}\boldsymbol{S}^{-1})\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y} \\ &+ 4a'(\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y})(\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y}+1) + 2a\operatorname{tr}\boldsymbol{S}^{-1}].\end{aligned} \tag{3.10}$$

If we substitute a and a' with $c_0 = k - p - 1$ and 0 respectively, we have

$$\begin{aligned}R^*(\widehat{\boldsymbol{\eta}}^{(u)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) = E[&(c_0^2 - 2(k-p-2)c_0)\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y} \\ &+ 2c_0(\operatorname{tr}\boldsymbol{S}^{-1})\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y} + 2c_0\operatorname{tr}\boldsymbol{S}^{-1}].\end{aligned} \tag{3.11}$$

The risk difference between $\widehat{\boldsymbol{\eta}}^{(1)}$ and $\widehat{\boldsymbol{\eta}}^{(u)}$ is given by

$$\begin{aligned}&R^*(\widehat{\boldsymbol{\eta}}^{(1)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) - R^*(\widehat{\boldsymbol{\eta}}^{(u)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) \\ &\quad= E[\{(a^2 - 2(k-p-2)a) - (c_0^2 - 2(k-p-2)c_0)\}\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y} \\ &\qquad + 2(a - c_0)(\operatorname{tr}\boldsymbol{S}^{-1})\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y} \\ &\qquad + 4a'(\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y})(\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y}+1) + 2(a-c_0)\operatorname{tr}\boldsymbol{S}^{-1}].\end{aligned}$$

Since $a < c_0$ and $(\operatorname{tr}\boldsymbol{S}^{-1})\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y} \geq \boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y}$, we have

$$\begin{aligned}&R(\widehat{\boldsymbol{\eta}}^{(1)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\eta}}^{(u)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) \\ &\quad\leq E[\{(a^2 - 2(k-p-3)a) - (c_0^2 - 2(k-p-3)c_0)\}\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y} \\ &\qquad\quad + 4a'(\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y})(\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y}+1) + 2(a-c_0)\operatorname{tr}\boldsymbol{S}^{-1}].\end{aligned}$$

If $c_0 > a \geq k - p - 5$ and $a' \leq 0$, then the following inequalities hold almost surely;

$$\begin{aligned}&\{(a^2 - 2(k-p-3)a) - (c_0^2 - 2(k-p-3)c_0)\}\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y} \leq 0, \\ &2(a - c_0)\operatorname{tr}\boldsymbol{S}^{-1} < 0\end{aligned}$$

and

$$4a'(\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y})(\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y}+1) \leq 0.$$

Obviously $R(\widehat{\boldsymbol{\eta}}^{(1)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) < R(\widehat{\boldsymbol{\eta}}^{(u)}; \boldsymbol{\theta}, \boldsymbol{\Sigma})$. □

The estimator (3.6), including $\widehat{\boldsymbol{\eta}}^{(u)}$ as a special case, is invariant w.r.t. the transformation with any nonsingular matrix $\boldsymbol{B}$,

$$\begin{aligned}&(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \to (\boldsymbol{B\theta}, \boldsymbol{B\Sigma B}^t), \\ &\boldsymbol{\eta} \to (\boldsymbol{B}^t)^{-1}\boldsymbol{\eta}, \\ &(\boldsymbol{Y}, \boldsymbol{S}) \to (\boldsymbol{BY}, \boldsymbol{BSB}^t),\end{aligned}$$

i.e.,

$$\widehat{\boldsymbol{\eta}}^{(1)}(\boldsymbol{BY},\boldsymbol{BSB}^t) = (\boldsymbol{B}^t)^{-1}\widehat{\boldsymbol{\eta}}^{(1)}(\boldsymbol{Y},\boldsymbol{S}).$$

The loss function (1.2) is invariant. Therefore on the risk comparison w.r.t. the loss function (1.2), Theorem 3.1 leads to the following inequality.

$$\begin{aligned}
&E[(\widehat{\boldsymbol{\eta}}^{(1)} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta})^t\boldsymbol{\Sigma}(\widehat{\boldsymbol{\eta}}^{(1)} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}) \mid \boldsymbol{\theta},\boldsymbol{\Sigma}]\\
&= E[(\widehat{\boldsymbol{\eta}}^{(1)} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta})^t(\widehat{\boldsymbol{\eta}}^{(1)} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}) \mid \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta},\boldsymbol{I}_p]\\
&= R(\widehat{\boldsymbol{\eta}}^{(1)};\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta},\boldsymbol{I}_p)\\
&\le R(\widehat{\boldsymbol{\eta}}^{(u)};\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta},\boldsymbol{I}_p)\\
&= E[(\widehat{\boldsymbol{\eta}}^{(u)} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta})^t(\widehat{\boldsymbol{\eta}}^{(u)} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}) \mid \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta},\boldsymbol{I}_p]\\
&= E[(\widehat{\boldsymbol{\eta}}^{(u)} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta})^t\boldsymbol{\Sigma}(\widehat{\boldsymbol{\eta}}^{(u)} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}) \mid \boldsymbol{\theta},\boldsymbol{\Sigma}].
\end{aligned}$$

We state this result as a corollary.

COROLLARY 3.1. *Suppose all the conditions of Theorem* 3.1 *are satisfied. Then* $\widehat{\boldsymbol{\eta}}^{(1)}$ *dominates* $\widehat{\boldsymbol{\eta}}^{(u)}$ *w.r.t. the loss function* (1.2).

We conclude this section by giving an example of $a(z)$ that satisfies the conditions of Theorem 3.1.

$$a(z) = k - p - 5 + \frac{4\alpha}{z+\alpha}, \qquad \alpha > 0. \tag{3.12}$$

4. New estimator II

In this section we present another type of estimators which is given by

$$\begin{aligned}
\widehat{\boldsymbol{\eta}}^{(2)} &= (\widehat{\eta}_1^{(2)},\ldots,\widehat{\eta}_p^{(2)})^t\\
&= \widehat{\boldsymbol{\eta}}^{(u)} - b(w)\boldsymbol{Y}\\
&= ((k-p-1)\boldsymbol{S}^{-1} - b(w)\boldsymbol{I}_p)\boldsymbol{Y},
\end{aligned} \tag{4.1}$$

where

$$w = \boldsymbol{Y}^t\boldsymbol{Y}.$$

(See e.g., Gupta and Nagar (1999) for the distribution of w and z in Section 3.)

As in the next theorem, this estimator also dominates $\widehat{\boldsymbol{\eta}}^{(u)}$.

THEOREM 4.1. *Suppose that* $b(w)$ *and its derivative* $b'(w)$ *are bounded continuous functions of* w *and* $b^2(w)w - 4b'(w)w - 2pb(w) < 0$, $\forall w > 0$, *then* $\widehat{\boldsymbol{\eta}}^{(2)}$ *dominates* $\widehat{\boldsymbol{\eta}}^{(u)}$ *w.r.t. the loss function* (1.1).

PROOF. $b(w)$ and $b'(w)$ being bounded and continuous, $\widehat{\boldsymbol{\eta}}^{(2)}$ in (4.1) satisfies all the conditions (1) to (5) (we omit the proof). From (2.8) and the fact

$$\begin{aligned}
\boldsymbol{D}\widehat{\boldsymbol{\eta}}^{(2)} &= (k-p-1)(\boldsymbol{D}\boldsymbol{S}^{-1})\boldsymbol{Y}\\
&= -\frac{1}{2}(k-p-1)((\operatorname{tr}\boldsymbol{S}^{-1})\boldsymbol{S}^{-1} + \boldsymbol{S}^{-2})\boldsymbol{Y}
\end{aligned}$$

and

$$\sum_{i=1}^{p} \frac{\partial \widehat{\eta}_i^{(2)}}{\partial y_i} = (k-p-1)\operatorname{tr} \boldsymbol{S}^{-1} - bp - 2b'w,$$

we have

$$\begin{aligned} R^*(\widehat{\boldsymbol{\eta}}^{(2)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) = E[&c_0(2-c_0)\boldsymbol{Y}^t\boldsymbol{S}^{-2}\boldsymbol{Y} + 2c_0(\operatorname{tr}\boldsymbol{S}^{-1})\boldsymbol{Y}^t\boldsymbol{S}^{-1}\boldsymbol{Y} + 2c_0\operatorname{tr}\boldsymbol{S}^{-1} \\ &+ b^2w - 2pb - 4b'w]. \end{aligned}$$

The risk difference between $\widehat{\boldsymbol{\eta}}^{(2)}$ and $\widehat{\boldsymbol{\eta}}^{(u)}$ is given by

$$R(\widehat{\boldsymbol{\eta}}^{(2)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\eta}}^{(u)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) = E[b^2w - 2pb - 4b'w].$$

If $b^2(w)w - 4b'(w)w - 2pb(w) < 0$, $\forall w > 0$, then $R(\widehat{\boldsymbol{\eta}}^{(2)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) < R(\widehat{\boldsymbol{\eta}}^{(u)}; \boldsymbol{\theta}, \boldsymbol{\Sigma})$. □

We give an example of $b(w)$ which satisfies the inequality $b^2(w)w - 4b'(w)w - 2pb(w) < 0$, $\forall w > 0$. Let

$$b(w) = \frac{\beta}{w+\alpha}, \qquad \alpha > 0, \beta > 0. \tag{4.2}$$

Then both $b(w)$ and $b'(w) = -\beta/(w+\alpha)^2$ are bounded continuous functions. Besides

$$b^2(w)w - 4b'(w)w - 2pb(w) = (w+\alpha)^{-2}(\beta(\beta+4-2p)w - 2p\alpha\beta) < 0$$

if $\beta < 2p - 4$. We have the following result.

COROLLARY 4.1. *Suppose $p \geq 3$. If $b(w)$ in* (4.1) *is given by* (4.2) *with $0 < \beta < 2p-4$, $0 \leq \alpha$, then $\widehat{\boldsymbol{\eta}}^{(2)}$ dominates $\widehat{\boldsymbol{\eta}}^{(u)}$ w.r.t. the loss function* (1.1).

PROOF. In the case $\alpha > 0$, it is obvious from Theorem 4.1. If we consider the limit value of $R(\widehat{\boldsymbol{\eta}}^{(2)}; \boldsymbol{\theta}, \boldsymbol{\Sigma})$ as α goes to zero, the case $\alpha = 0$ can be proved as well. □

Note that if $\beta < p - 2(\equiv \beta^*)$, then $\widehat{\boldsymbol{\eta}}^{(2)}$ defined by (4.1) and (4.2) is dominated by

$$\widehat{\boldsymbol{\eta}}^{(2*)} = (k-p-1)\boldsymbol{S}^{-1}\boldsymbol{Y} - \frac{p-2}{w+\alpha}\boldsymbol{Y} \tag{4.3}$$

which is a special case of $\widehat{\boldsymbol{\eta}}^{(2)}$ when β of (4.2) is given by β^*, since

$$\begin{aligned} &R(\widehat{\boldsymbol{\eta}}^{(2*)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\eta}}^{(2)}; \boldsymbol{\theta}, \boldsymbol{\Sigma}) \\ &\quad = E[(w+\alpha)^{-2}\{(\beta^*(\beta^*+4-2p) - \beta(\beta+4-2p))w - 2p\alpha(\beta^*-\beta)\}] \end{aligned}$$

and

$$\beta^*(\beta^*+4-2p) < \beta(\beta+4-2p).$$

Therefore the class of estimators given by (4.1) and (4.2) is naturally to be confined to the case $p - 2 \leq \beta < 2p - 4$ with respect to admissibility.

5. Monte Carlo Simulation

We carried out a small-scale Monte Carlo Simulation to compare the following estimators:

1. $\widehat{\eta}^{(u)}$ given by (1.3);
2. $\widehat{\eta}^{(H_1)}$ given by (1.4) with $a = k - p - 3$;

Table 1. Risk of estimators.

$v = 1,\ r = 0$			$v = 1,\ r = 0.99$			$v = 1,\ r = -0.4$		
$t = 0$	Risk	PRIAL	$t = 0$	Risk	PRIAL	$t = 0$	Risk	PRIAL
u	3.271	0	u	217.385	0	u	6.995	0
H_1	2.993	9	H_1	198.893	9	H_1	6.4	9
H_2	3.156	4	H_2	206.632	5	H_2	6.582	6
DS	2.897	11	DS	189.995	13	DS	6.151	12
1	3.214	2	1	213.477	2	1	6.872	2
2	2.223	32	2	208.468	4	2	5.705	18
$t = 1$	Risk	PRIAL	$t = 1$	Risk	PRIAL	$t = 1$	Risk	PRIAL
u	3.557	0	u	216.827	0	u	10.812	0
H_1	3.262	8	H_1	198.383	9	H_1	10.06	7
H_2	3.428	4	H_2	206.152	5	H_2	10.193	6
DS	3.161	11	DS	189.508	13	DS	9.83	9
1	3.443	3	1	212.442	2	1	9.92	8
2	3.144	12	2	212.481	2	2	10.61	2
$t = 10$	Risk	PRIAL	$t = 10$	Risk	PRIAL	$t = 10$	Risk	PRIAL
u	30	0	u	6.73×10^2	0	u	397.103	0
H_1	28.04	7	H_1	6.16×10^2	8	H_1	379.912	4
H_2	28.401	5	H_2	6.40×10^2	5	H_2	375.413	5
DS	27.362	9	DS	5.88×10^2	13	DS	381.137	4
1	27.106	10	1	5.93×10^2	12	1	389.153	2
2	29.998	0	2	6.73×10^2	0	2	397.137	0
$t = 100$	Risk	PRIAL	$t = 100$	Risk	PRIAL	$t = 100$	Risk	PRIAL
u	2.67×10^3	0	u	4.42×10^4	0	u	3.88×10^4	0
H_1	2.49×10^3	7	H_1	4.05×10^4	8	H_1	3.69×10^4	5
H_2	2.51×10^3	6	H_2	4.21×10^4	5	H_2	3.65×10^4	6
DS	2.43×10^3	9	DS	3.87×10^4	13	DS	3.70×10^4	5
1	2.44×10^3	8	1	3.69×10^4	17	1	3.80×10^4	2
2	2.67×10^3	0	2	4.42×10^4	0	2	3.88×10^4	0
$t = 10000$	Risk	PRIAL	$t = 10000$	Risk	PRIAL	$t = 10000$	Risk	PRIAL
u	2.69×10^7	0	u	4.54×10^8	0	u	3.88×10^8	0
H_1	2.52×10^7	6	H_1	4.15×10^8	8	H_1	3.67×10^8	5
H_2	2.54×10^7	6	H_2	4.31×10^8	5	H_2	3.63×10^8	6
DS	2.46×10^7	8	DS	3.97×10^8	13	DS	3.67×10^8	5
1	2.48×10^7	8	1	3.79×10^8	17	1	3.77×10^8	3
2	2.69×10^7	0	2	4.54×10^8	0	2	3.88×10^8	0

3. $\widehat{\boldsymbol{\eta}}^{(H_2)}$ given by (1.5) with $t(u) = 2/(k-p-5)$;
4. $\widehat{\boldsymbol{\eta}}^{(DS)}$ given by (1.6) with $a = k-p-4$, $b = 1$;
5. $\widehat{\boldsymbol{\eta}}^{(1)}$ given by (3.6) and (3.12) with $\alpha = 1$;
6. $\widehat{\boldsymbol{\eta}}^{(2)}$ given by (4.1) and (4.2) with $\alpha = 0$, $\beta = p-2$.

For $p = 3$, $k = 50$,

$$\boldsymbol{\theta} = (t,t,t)^t, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} v & r & r \\ r & v & r \\ r & r & v \end{pmatrix}$$

with various values of t, r, v, we generated 5000 $\boldsymbol{Y}$'s and 5000 $\boldsymbol{S}$'s independently according to the following distributions:

$$\boldsymbol{Y} \sim N_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}), \qquad \boldsymbol{S} \sim \boldsymbol{W}_p(k, \boldsymbol{\Sigma}).$$

Table 1 shows simulated risk and PRIAL of each estimator $\widehat{\boldsymbol{\eta}}^{(\cdot)}$. PRIAL defined by

$$\frac{\text{the risk of } \widehat{\boldsymbol{\eta}}^{(u)} - \text{the risk of } \widehat{\boldsymbol{\eta}}^{(\cdot)}}{\text{the risk of } \widehat{\boldsymbol{\eta}}^{(u)}} \times 100$$

shows the risk reduction relative to $\widehat{\boldsymbol{\eta}}^{(u)}$. Note that PRIAL is rounded into the nearest integer.

We notice that

1. $\widehat{\boldsymbol{\eta}}^{(1)}$ records PRIAL up to 17%, while the maximum PRIAL of $\widehat{\boldsymbol{\eta}}^{(2)}$ reaches as high as 32%.
2. There is no dominance relationship between the estimators other than $\widehat{\boldsymbol{\eta}}^{(u)}$.

Acknowledgements

We appreciate valuable advice of Professor Y. Fujikoshi, which motivated us to consider this problem. We also thank Dr. Sanat K. Sarkar for his kindness to send us a copy of his unpublished joint work. Thanks go to Professor F. Funaoka for valuable advice on the application of the discriminant analysis. We would like to thank the referees for their valuable comments which lead to the improvement of the paper. The first author is deeply grateful to the Department of Mathematics and Statistics, Bowling Green State University for giving him a pleasant research environment.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, **23**, 589–609.

Altman, E. I. (1993). *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress and Profiting from Bankruptcy*, 2nd ed., John Wiley and Sons, New York.

Dey, D. K. and Srinivasan, C. (1991). On estimation of discriminant coefficients, *Statistics and Probability Letters*, **11**, 189–193.

Grice, J. S. and Ingram, R. W. (2001). Tests of the generalizability of Altman's bankruptcy prediction model, *Journal of Business Research*, **54**, 53–61.

Gupta, A. K. and Nagar, D. K. (1999). *Matrix Variate Distributions*, Chapman & Hall/CRC, Boca Raton.

Haff, L. R. (1982). Identities for the inverse Wishart distribution with computational results in linear and quadratic discrimination, *Sankhyā Series B*, **44**, 245–258.

Haff, L. R. (1986). On linear log-odds and estimation of discriminant coefficients, *Communications and Statistics, Theory and Method*, **15**, 2131–2144.

Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, 4th ed., Prentice-Hall, Upper Saddle River.

Rukhin, A. L. (1992). Generalized Bayes estimators of a normal discriminant function, *Journal of Multivariate Analysis*, **41**, 154–162.

Sarkar, S. K. and Krishnamoorthy, K. (1991). Estimation of a function of multivariate normal parameters, Tech. Report, 91-1, Department of Statistics, Temple University, Pennsylvania.