Ann. Inst. Statist. Math.
Vol. 55, No. 3, 447–466 (2003)

ON ESTIMATION IN MULTIVARIATE LINEAR CALIBRATION WITH ELLIPTICAL ERRORS

HISAYUKI TSUKUMA

The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan,
e-mail: tsukuma@ism.ac.jp

(Received September 30, 2002)

Abstract. The estimation problem in multivariate linear calibration with elliptical errors is considered under a loss function which can be derived from the Kullback-Leibler distance. First, we discuss the problem under normal errors and give unbiased estimate of risk of an alternative estimator by means of the Stein and Stein-Haff identities for multivariate normal distribution. From the unbiased estimate of risk, it is shown that a shrinkage estimator improves on the classical estimator under the loss function. Furthermore, from the extended Stein and Stein-Haff identities for our elliptically contoured distribution, the above result under normal errors is extended to the estimation problem under elliptical errors. We show that the shrinkage estimator obtained under normal models is better than the classical estimator under elliptical errors with the above loss function and hence we establish the robustness of the above shrinkage estimator.

Key words and phrases: Elliptically contoured distribution, Kullback-Leibler distance, multivariate linear model, shrinkage estimator.

1. Introduction

The calibration problem occurs in measurement settings where two measurement methods are available: One is extremely accurate but expensive (or time-consuming) while the other is less accurate but easier and fast. The functional relation between the two types of measurements is assessed through a calibration experiment, where the values of both measurements are known; this relation is then used in subsequent experiments to predict the value of the more precise measurement based on a sample of the more approximate measurement.

This setting is of major importance in physical and chemical measurements and we refer the reader to Rosenblatt and Spiegelman (1981) for a general discussion on the practical issues of calibration. For a detailed and recent survey of the calibration problem, see Brown (1982, 1993), Osborne (1991), and Sundberg (1999).

In this paper we consider the multivariate linear calibration model. Let $\boldsymbol{Y}$ and $\boldsymbol{Y}_0$ be, respectively, $n \times p$ and $m \times p$ random matrices of response variables and also let $\boldsymbol{X}$ be an $n \times q$ matrix of explanatory variables with full rank. Consider the calibration experiment and the prediction experiment which can be represented as, respectively,

$$\boldsymbol{Y} = \boldsymbol{1}_n \boldsymbol{\alpha}^t + \boldsymbol{X}\boldsymbol{\Theta} + \boldsymbol{\epsilon}, \tag{1.1}$$

$$\boldsymbol{Y}_0 = \boldsymbol{1}_m \boldsymbol{\alpha}^t + \boldsymbol{1}_m \boldsymbol{x}_0^t \boldsymbol{\Theta} + \boldsymbol{\epsilon}_0, \tag{1.2}$$

where $\mathbf{1}_l$ is the $l \times 1$ vector consisting of ones, $\boldsymbol{\alpha}$ and $\boldsymbol{\Theta}$ are, respectively, $p \times 1$ vector and $q \times p$ matrix of unknown parameters, and $\boldsymbol{x}_0$ is the $q \times 1$ vector to predict. Here, we denote by $\boldsymbol{A}^t$ the transpose of a matrix $\boldsymbol{A}$. Furthermore $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_0$ are, respectively, $n \times p$ and $m \times p$ error matrices with mean zero matrices. We assume that $p \geq q$, $n+m-q-2 \geq p$, and $\boldsymbol{X}^t\mathbf{1}_n = \mathbf{0}_{q\times 1}$. Our problem is to predict $\boldsymbol{x}_0$ based on $(\boldsymbol{Y}, \boldsymbol{X})$ and $\boldsymbol{Y}_0$.

We assume two cases of error distributions: (I) The rows of the error matrices, $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_0$, are independently and identically distributed as the p-variate normal distributions with mean zero vector and covariance matrix $\boldsymbol{\Sigma}$, abbreviated by $\mathcal{N}_p(\mathbf{0}_{p\times 1}, \boldsymbol{\Sigma})$. (II) The error matrices, $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_0$, are jointly distributed as the elliptically contoured distribution with its density function

$$|\boldsymbol{\Sigma}|^{-(n+m)/2} f(\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}^t\boldsymbol{\epsilon} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}_0^t\boldsymbol{\epsilon}_0\}), \tag{1.3}$$

where f is an unknown, nonnegative function on $[0, \infty)$ and $\boldsymbol{\Sigma}$ is a $p \times p$ scale matrix. In both cases (I) and (II), we assume that $\boldsymbol{\Sigma}$ is unknown and positive-definite. Here, we denote by $\mathrm{tr}(\boldsymbol{A})$ and $|\boldsymbol{A}|$ the trace and the determinant of a squared matrix $\boldsymbol{A}$.

There has been plenty of literature on the problem of estimating $\boldsymbol{x}_0$ in (1.2) under the errors (I). Then two estimators are well-known; one is the *classical estimator* and the other is the *inverse regression estimator*. Now, denote the least squares estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\Theta}$ by

$$\hat{\boldsymbol{\alpha}} = \bar{\boldsymbol{y}}, \qquad \widehat{\boldsymbol{\Theta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{Y}, \tag{1.4}$$

where $\bar{\boldsymbol{y}} = \boldsymbol{Y}^t\mathbf{1}_n/n$. Let

$$\begin{aligned} &\bar{\boldsymbol{y}}_0 = \boldsymbol{Y}_0^t\mathbf{1}_m/m, \qquad \boldsymbol{V}_0 = (\boldsymbol{Y}_0 - \mathbf{1}_m\bar{\boldsymbol{y}}_0^t)^t(\boldsymbol{Y}_0 - \mathbf{1}_m\bar{\boldsymbol{y}}_0^t), \\ &\boldsymbol{V} = (\boldsymbol{Y} - \mathbf{1}_n\hat{\boldsymbol{\alpha}}^t - \boldsymbol{X}\widehat{\boldsymbol{\Theta}})^t(\boldsymbol{Y} - \mathbf{1}_n\hat{\boldsymbol{\alpha}}^t - \boldsymbol{X}\widehat{\boldsymbol{\Theta}}), \qquad \text{and} \qquad \boldsymbol{S} = \boldsymbol{V} + \boldsymbol{V}_0. \end{aligned} \tag{1.5}$$

Brown (1982) derived the classical and the inverse regression estimators which are given by, respectively,

$$\hat{\boldsymbol{x}}_0 = (\widehat{\boldsymbol{\Theta}}\boldsymbol{S}^{-1}\widehat{\boldsymbol{\Theta}}^t)^{-1}\widehat{\boldsymbol{\Theta}}\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}_0 - \bar{\boldsymbol{y}}) \tag{1.6}$$

and

$$\check{\boldsymbol{x}}_0 = \{(\boldsymbol{X}^t\boldsymbol{X})^{-1} + \widehat{\boldsymbol{\Theta}}\boldsymbol{V}^{-1}\widehat{\boldsymbol{\Theta}}^t\}^{-1}\widehat{\boldsymbol{\Theta}}\boldsymbol{V}^{-1}(\bar{\boldsymbol{y}}_0 - \bar{\boldsymbol{y}}). \tag{1.7}$$

The classical estimator (1.6) is the restricted maximum likelihood estimator and for $n \to \infty$ and $m \to \infty$ it is consistent when $\boldsymbol{\Theta} \neq \mathbf{0}$ but the inverse regression estimator (1.7) is not consistent. For details of comparison between the classical and the inverse regression estimators see, for example, Brown (1982, 1993).

The main interest of this paper is an improvement on the classical estimator (1.6) from a decision-theoretic point of view. When $q = 1$, $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_p$ and σ^2 is unknown in models (1.1) and (1.2), Kubokawa and Robert (1994) showed, under the squared loss, that the classical estimator is inadmissible and that the inverse regression estimator is admissible. Srivastava (1995) showed the inadmissibility of the classical estimator and the admissibility of the inverse regression estimator when $q = 1$ and $\boldsymbol{\Sigma}$ is fully unknown. Furthermore, when $q > 1$ in (1.1) and (1.2), Tsukuma (2002) discussed the problem of estimating $\boldsymbol{x}_0$ under the quadratic loss function

$$L_0(\tilde{\boldsymbol{x}}_0; \boldsymbol{x}_0) = (1/c_{n,m})(\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0)^t(\boldsymbol{X}^t\boldsymbol{X})^{-1}(\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0), \tag{1.8}$$

where $\tilde{x}_0$ is an estimator of $\boldsymbol{x}_0$ and $c_{n,m} = 1/n + 1/m$. Tsukuma (2002) proposed an alternative estimator over the classical estimator and showed that the inverse regression estimator is admissible under the loss (1.8). On the other hand Branco *et al.* (2000) treated a Bayesian analysis of the calibration problem under the multivariate linear model with elliptical errors whose density is different from (1.3) and they showed that a Bayes estimator for a noninformative prior is the inverse regression estimator.

In this paper we discuss the problem of estimating $\boldsymbol{x}_0$ under the *quasi*-loss function

$$L(\tilde{\boldsymbol{x}}_0; \boldsymbol{x}_0) = (1/c_{n,m})(\widehat{\boldsymbol{\Theta}}^t\tilde{\boldsymbol{x}}_0 - \boldsymbol{\Theta}^t\boldsymbol{x}_0)^t\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^t\tilde{\boldsymbol{x}}_0 - \boldsymbol{\Theta}^t\boldsymbol{x}_0). \tag{1.9}$$

Then the accuracy of an estimator $\tilde{x}_0$ is measured by the risk function $R(\tilde{\boldsymbol{x}}_0; \boldsymbol{x}_0) = E[L(\tilde{\boldsymbol{x}}_0; \boldsymbol{x}_0)]$. The loss function L can be regarded as a quadratic loss function in the problem of estimating $\boldsymbol{\Theta}^t\boldsymbol{x}_0$ by an estimator $\widehat{\boldsymbol{\Theta}}^t\tilde{\boldsymbol{x}}_0$ but L is not a loss function in terms of $\boldsymbol{x}_0$ and $\tilde{\boldsymbol{x}}_0$. The usage of L is motivated by the following reasons: (1) If $\boldsymbol{\alpha}$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$ are known under normal errors, then the maximum likelihood estimator is $\hat{\boldsymbol{x}}_0^{ML} = (\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Theta}^t)^{-1}\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{y}}_0 - \boldsymbol{\alpha})$ and $\hat{\boldsymbol{x}}_0^{ML} \sim \mathcal{N}_q(\boldsymbol{x}_0, (\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Theta}^t)^{-1})$. Thus it seems that the behavior of L is similar to that of a natural loss function

$$L_1(\tilde{\boldsymbol{x}}_0; \boldsymbol{x}_0) = (1/c_{n,m})(\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0)^t\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Theta}^t(\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0). \tag{1.10}$$

(2) Under normal errors, the loss function L can be derived from the Kullback-Leibler distance

$$\int \left\{ \log \frac{p(\bar{\boldsymbol{y}}, \widehat{\boldsymbol{\Theta}}, \boldsymbol{S}, \bar{\boldsymbol{y}}_0 \mid \tilde{\alpha}, \widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\Sigma}}, \tilde{\boldsymbol{x}}_0)}{p(\bar{\boldsymbol{y}}, \widehat{\boldsymbol{\Theta}}, \boldsymbol{S}, \bar{\boldsymbol{y}}_0 \mid \boldsymbol{\alpha}, \boldsymbol{\Theta}, \boldsymbol{\Sigma}, \boldsymbol{x}_0)} \right\} p(\bar{\boldsymbol{y}}, \widehat{\boldsymbol{\Theta}}, \boldsymbol{S}, \bar{\boldsymbol{y}}_0 \mid \tilde{\alpha}, \widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\Sigma}}, \tilde{\boldsymbol{x}}_0) d\bar{\boldsymbol{y}} d\widehat{\boldsymbol{\Theta}} d\boldsymbol{S} d\bar{\boldsymbol{y}}_0,$$

where $p(\bar{\boldsymbol{y}}, \widehat{\boldsymbol{\Theta}}, \boldsymbol{S}, \bar{\boldsymbol{y}}_0 \mid \boldsymbol{\alpha}, \boldsymbol{\Theta}, \boldsymbol{\Sigma}, \boldsymbol{x}_0)$ denotes a joint density function of $(\bar{\boldsymbol{y}}, \widehat{\boldsymbol{\Theta}}, \boldsymbol{S}, \bar{\boldsymbol{y}}_0)$. Here $(\bar{\boldsymbol{y}}, \widehat{\boldsymbol{\Theta}}, \boldsymbol{V}, \bar{\boldsymbol{y}}_0)$ is given by (1.4) and (1.5) and $(\tilde{\alpha}, \widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\Sigma}}, \tilde{\boldsymbol{x}}_0)$ is an estimator of $(\boldsymbol{\alpha}, \boldsymbol{\Theta}, \boldsymbol{\Sigma}, \boldsymbol{x}_0)$.

This paper is organized in the following manner: In Section 2, the problem of estimating $\boldsymbol{x}_0$ is considered under the errors (I), i.e., the rows of the error matrices are mutually and independently distributed as the multivariate normal distributions. First we derive a canonical form for this setup and give unbiased estimate of risk of an alternative estimator via the Stein and Stein-Haff identities for multivariate normal distribution. From this unbiased estimate of risk, it is shown that shrinkage estimators improve on the classical estimator (1.6) under the loss function L. For example, one of the shrinkage estimators is the James-Stein type estimator (see James and Stein (1961))

$$\hat{\boldsymbol{x}}_0^{JS} = \left(1 - \frac{c_{n,m}(q-2)}{(n+m-q-p+1)(\bar{\boldsymbol{y}}_0 - \bar{\boldsymbol{y}})^t\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}_0 - \bar{\boldsymbol{y}})}\right)\hat{\boldsymbol{x}}_0, \quad \text{for} \quad q \geq 3,$$

which is different from improved estimators given by Kubokawa and Robert (1994) and Tsukuma (2002). Next, in Section 3 we discuss the problem with the errors (II), i.e., the error matrices are jointly and uncorrelatedly distributed as an elliptically contoured distribution. From the extended Stein and Stein-Haff identities for our elliptically contoured distribution due to Kubokawa and Srivastava (1999, 2001), the above domination under normal errors is extended to the estimation problem under elliptical errors. Monte Carlo simulations in special case of an elliptical distribution is carried out to evaluate the risk performance under the loss function L_1 since it is very difficult to prove the improvement under the loss function L_1. From this simulations, we illustrate that a shrinkage estimator is better than the classical estimator even if the loss function L_1 is used. Furthermore, since the problem with the errors (II) is not independent sampling, we also conduct a simulation study based on independently and identically sampling model from an elliptically contoured distribution. Under this setup, we also show that the James-Stein type estimator is numerically better than the classical estimator under the loss function L_1. Finally, in Section 4 we state some technical lemmas and give proofs of theorems in Sections 2 and 3.

2. Improving on the classical estimator under normal errors

In this section, we consider an improvement on the classical estimator under normal errors. First, we give a canonical form of this problem and, next, state main theorems of this section. Proofs of theorems and corollaries are postponed to Subsection 4.1.

2.1 *A canonical form*

We first define the following notation. The Kronecker product of matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ is denoted by "$\boldsymbol{A} \otimes \boldsymbol{C}$". For any $q \times p$ matrix $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_q)^t$ with $p \times 1$ vectors $\boldsymbol{z}_i$, we write $\mathrm{vec}(\boldsymbol{Z}^t) = (\boldsymbol{z}_1^t, \ldots, \boldsymbol{z}_q^t)^t$. '$\boldsymbol{Z} \sim \mathcal{N}_{q\times p}(\boldsymbol{M}, \boldsymbol{A} \otimes \boldsymbol{C})$' indicates that $\mathrm{vec}(\boldsymbol{Z}^t)$ follows multivariate normal distribution with mean $\mathrm{vec}(\boldsymbol{M}^t)$ and covariance matrix $\boldsymbol{A} \otimes \boldsymbol{C}$. Furthermore, '$\mathcal{W}_p(\boldsymbol{\Sigma}, k)$' stands for the Wishart distribution with degrees of freedom k and mean $k\boldsymbol{\Sigma}$.

The classical estimator for unknown $\boldsymbol{x}_0$ is rewritten as

$$\hat{\boldsymbol{x}}_0 = (\widehat{\boldsymbol{\Theta}} \boldsymbol{S}^{-1} \widehat{\boldsymbol{\Theta}}^t)^{-1} \widehat{\boldsymbol{\Theta}} \boldsymbol{S}^{-1} (\bar{\boldsymbol{y}}_0 - \bar{\boldsymbol{y}}), \tag{2.1}$$

where $\bar{\boldsymbol{y}}$, $\widehat{\boldsymbol{\Theta}}$, $\boldsymbol{S}$, and $\bar{\boldsymbol{y}}_0$ are given in (1.4) and (1.5). We here note that these statistics $\bar{\boldsymbol{y}}$, $\widehat{\boldsymbol{\Theta}}$, $\boldsymbol{S}$, and $\bar{\boldsymbol{y}}_0$ are mutually and independently distributed as

$$\begin{aligned} &\bar{\boldsymbol{y}} \sim \mathcal{N}_p(\boldsymbol{\alpha}, (1/n)\boldsymbol{\Sigma}), \qquad \widehat{\boldsymbol{\Theta}} \sim \mathcal{N}_{q\times p}(\boldsymbol{\Theta}, (\boldsymbol{X}^t\boldsymbol{X})^{-1} \otimes \boldsymbol{\Sigma}), \\ &\boldsymbol{S} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, l), \qquad \text{and} \qquad \bar{\boldsymbol{y}}_0 \sim \mathcal{N}_p(\boldsymbol{\alpha} + \boldsymbol{\Theta}^t \boldsymbol{x}_0, (1/m)\boldsymbol{\Sigma}) \end{aligned}$$

for $l = n + m - q - 2 \geq p$.

Let $c_{n,m} = 1/n + 1/m$, $\boldsymbol{z} = c_{n,m}^{-1/2}(\bar{\boldsymbol{y}}_0 - \bar{\boldsymbol{y}})$, and $\boldsymbol{B} = (\boldsymbol{X}^t\boldsymbol{X})^{1/2}\widehat{\boldsymbol{\Theta}}$. Here, we denote by $\boldsymbol{A}^{1/2}$ a symmetric matrix such that $\boldsymbol{A} = \boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2}$. Then $\boldsymbol{B}$, $\boldsymbol{S}$, and $\boldsymbol{z}$ are mutually and independently distributed as

$$\boldsymbol{B} \sim \mathcal{N}_{q\times p}(\boldsymbol{\beta}, \boldsymbol{I}_q \otimes \boldsymbol{\Sigma}), \qquad \boldsymbol{S} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, l), \qquad \text{and} \qquad \boldsymbol{z} \sim \mathcal{N}_p(\boldsymbol{\beta}^t\boldsymbol{\xi}, \boldsymbol{\Sigma}), \tag{2.2}$$

where $\boldsymbol{\beta} = (\boldsymbol{X}^t\boldsymbol{X})^{1/2}\boldsymbol{\Theta}$ and $\boldsymbol{\xi} = c_{n,m}^{-1/2}(\boldsymbol{X}^t\boldsymbol{X})^{-1/2}\boldsymbol{x}_0$. The loss function (1.9) can be written as

$$L(\tilde{\boldsymbol{\xi}}; \boldsymbol{\xi}) = (\boldsymbol{B}^t\tilde{\boldsymbol{\xi}} - \boldsymbol{\beta}^t\boldsymbol{\xi})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{B}^t\tilde{\boldsymbol{\xi}} - \boldsymbol{\beta}^t\boldsymbol{\xi}). \tag{2.3}$$

To express the classical estimator (2.1) with $\boldsymbol{B}$, $\boldsymbol{S}$ and $\boldsymbol{z}$, we put $\hat{\boldsymbol{\xi}} = c_{n,m}^{-1/2}(\boldsymbol{X}^t\boldsymbol{X})^{-1/2}\hat{\boldsymbol{x}}_0$ to have

$$\hat{\boldsymbol{\xi}} = (\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}. \tag{2.4}$$

Similarly, using the statistics $\boldsymbol{B}$ and $\boldsymbol{z}$, we can write the inverse regression estimator (1.7) as

$$\check{\boldsymbol{\xi}} = (\boldsymbol{I}_q + \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{z}, \tag{2.5}$$

where $\check{\boldsymbol{\xi}} = c_{n,m}^{-1/2}(\boldsymbol{X}^t\boldsymbol{X})^{-1/2}\check{\boldsymbol{x}}_0$. We here note that $\boldsymbol{V} \sim \mathcal{W}_p(\Sigma, l_1)$ where $l_1 = n - q - 1$ and that the statistics $\boldsymbol{V}$, $\boldsymbol{B}$, and $\boldsymbol{z}$ are mutually independent.

In next subsection we treat the calibration problem on the model (2.2) and discuss an improvement on the classical estimator (2.4) under the loss (2.3).

2.2 *Improved estimator and unbiased estimate of its risk*

Note that the estimation problem on the model (2.2) is invariant under the group of transformations:

$$\begin{array}{lll} \beta \to \boldsymbol{Q}\beta\boldsymbol{P}, & \Sigma \to \boldsymbol{P}^t\Sigma\boldsymbol{P}, & \boldsymbol{\xi} \to \boldsymbol{Q}\boldsymbol{\xi}, \\ \boldsymbol{B} \to \boldsymbol{Q}\boldsymbol{B}\boldsymbol{P}, & \boldsymbol{S} \to \boldsymbol{P}^t\boldsymbol{S}\boldsymbol{P}, & \boldsymbol{z} \to \boldsymbol{P}^t\boldsymbol{z} \end{array}$$

for any $q \times q$ orthogonal matrix $\boldsymbol{Q}$ and any $p \times p$ nonsingular matrix $\boldsymbol{P}$.

Now, for estimating $\boldsymbol{\xi}$ in (2.2) under the loss (2.3), we consider a class of estimators

$$\hat{\boldsymbol{\xi}}(\phi, \boldsymbol{G}) = \phi\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}, \tag{2.6}$$

where ϕ is a scalar-valued function of $\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$ and $\boldsymbol{G}$ is a $q \times q$ symmetric matrix whose elements are functions of $\boldsymbol{F} = \boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t$. The estimators (2.6) can be interpreted as an extension of the classical estimator (2.4). Remark that for $m = 1$ the inverse regression estimator (2.5) belongs to the above class of estimators since $\boldsymbol{S} \equiv \boldsymbol{V}$ but the estimator (2.5) does not belong to it for $m \geq 2$.

From the Stein identity for the multivariate normal distribution and the Stein-Haff identity for the Wishart distribution, we can evaluate the risk of the estimators (2.6) as follows:

THEOREM 2.1. *Let statistics $\boldsymbol{B}$, $\boldsymbol{S}$, and $\boldsymbol{z}$ be defined as (2.2) and let $\boldsymbol{F} = \boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t$. Further, denote by D_F differential operator in terms of $\boldsymbol{F} = (F_{ij})$ where the (i,j)-element of D_F is $\{D_F\}_{ij} = (1/2)(1 + \delta_{ij})\partial/\partial F_{ij}$ with the Kronecker delta δ_{ij}. Suppose that we wish to estimate $\boldsymbol{\xi}$ in (2.2) by*

$$\hat{\boldsymbol{\xi}}(\phi, \boldsymbol{G}) = \phi\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z},$$

where ϕ is a scalar-valued function of $t = \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$ and $\boldsymbol{G} = (G_{ij})$ is a $q \times q$ symmetric matrix whose elements are functions of $\boldsymbol{F}$. Then, under the loss L given in (2.3), the risk of the estimators $\hat{\boldsymbol{\xi}}(\phi, \boldsymbol{G})$ can be represented as

$$\begin{aligned} R(\hat{\boldsymbol{\xi}}(\phi, \boldsymbol{G}), \boldsymbol{\xi}) = E[&-p + 4\phi'\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} + 2\phi\,\mathrm{tr}(\boldsymbol{F}\boldsymbol{G}) \\ &+ (l - p - 1)\,\mathrm{tr}[\boldsymbol{S}^{-1}(\boldsymbol{z} - \phi\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})(\boldsymbol{z} - \phi\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})^t] \\ &+ 4\phi'\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - \phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}) \\ &+ 4\phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{I}_q - \phi\boldsymbol{G}\boldsymbol{F})\{(\boldsymbol{F}D_F)^t\boldsymbol{G}\}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} \\ &+ 2\phi[\mathrm{tr}(\boldsymbol{F}\boldsymbol{G})](\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - \phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}) \\ &+ 2\phi(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} - \phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{F}\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})], \end{aligned} \tag{2.7}$$

where the expectation is taken with respect to (2.2) *and* $\phi' = d\phi/dt$. *Here* '$\{(\boldsymbol{F}D_F)^t\boldsymbol{G}\}$' *indicates that* D_F *acts only on* $\boldsymbol{G}$ *and that the* (i,j)*-element of* $\{(\boldsymbol{F}D_F)^t\boldsymbol{G}\}$ *is given by*

$$\{(\boldsymbol{F}D_F)^t\boldsymbol{G}\}_{ij} = \sum_{a,b} F_{ab}[\{D_F\}_{ia}G_{bj}].$$

Note that the content of expectation in the right-hand side of (2.7) is an unbiased estimate of risk in terms of the estimators $\hat{\boldsymbol{\xi}}(\phi, \boldsymbol{G})$. In Theorem 2.1, putting $\phi = 1$ and $\boldsymbol{G} = \boldsymbol{F}^{-1}$, we obtain unbiased estimate of risk of the classical estimator:

COROLLARY 2.1. *Under the loss* L, *the risk of the classical estimator* (2.4) *can be expressed as*

$$R(\hat{\boldsymbol{\xi}}, \boldsymbol{\xi}) = E[-p + 2q + (l - p - 1 + 2q)(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})]. \tag{2.8}$$

For improving on the classical estimator $\hat{\boldsymbol{\xi}}$, we consider shrinkage estimators (see Baranchik (1970))

$$\hat{\boldsymbol{\xi}}(\psi) = (1 - \psi/t)(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}, \tag{2.9}$$

where ψ is a differentiable function of $t = \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$. From Theorem 2.1 and Corollary 2.1, we establish the following dominance result:

THEOREM 2.2. *Assume that* $q \geq 3$. *If*
(i) ψ *is nondecreasing, and*
(ii) $0 \leq \psi \leq 2(q-2)/(l-p+3)$,
then the shrinkage estimators (2.9) *improve on the classical estimator* (2.4) *under the loss* L.

For example, one of the shrinkage estimators is the James-Stein type estimator (see James and Stein (1961))

$$\hat{\boldsymbol{\xi}}^{JS} = \left(1 - \frac{q-2}{(l-p+3)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}}\right)\hat{\boldsymbol{\xi}} \tag{2.10}$$

for $q \geq 3$.

Theorem 2.2 indicates that under the loss L the estimators $\hat{\boldsymbol{\xi}}(\psi)$ improve on the classical estimator $\hat{\boldsymbol{\xi}}$ by statistics $\boldsymbol{z}$ and $\boldsymbol{S}$. Since the statistic $\boldsymbol{z}$ has much information on $\boldsymbol{\xi}$, the result of Theorem 2.2 seems to be natural. On the other hand, Tsukuma (2002) proposed an improved estimator on the classical estimator under the loss function (1.8). The improved estimator is constructed by means of statistics $\boldsymbol{B}$ and $\boldsymbol{S}$ and hence it is different from the estimators $\hat{\boldsymbol{\xi}}(\psi)$. See also Kubokawa and Robert (1994).

Remark 2.1. In Theorem 2.1, we replace $\boldsymbol{S}$ by $\boldsymbol{V}$ and put $\phi = 1$ and $\boldsymbol{G} = (\boldsymbol{I}_q + \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^t)^{-1}$ to evaluate risk of the inverse regression estimator (2.5) as follows:

COROLLARY 2.2.

$$
\begin{aligned}
(2.11)\qquad R(\check{\boldsymbol{\xi}}, \boldsymbol{\xi}) &= E[L(\check{\boldsymbol{\xi}}; \boldsymbol{\xi})] \\
&= E[-p + 2\,\mathrm{tr}\{\boldsymbol{F}_1(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\} \\
&\quad + (l_1 - p - 1)\,\mathrm{tr}\{\boldsymbol{V}^{-1}(\boldsymbol{z} - \boldsymbol{B}^t(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{z}) \\
&\quad \times (\boldsymbol{z} - \boldsymbol{B}^t(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{z})^t\} \\
&\quad + 2\boldsymbol{z}^t\boldsymbol{V}^{-1}\boldsymbol{B}^t(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\boldsymbol{A}(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{z} \\
&\quad + 2(\mathrm{tr}\{\boldsymbol{F}_1(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\}) \\
&\quad \times (\boldsymbol{z}^t\boldsymbol{V}^{-1}\boldsymbol{z} - \boldsymbol{z}^t\boldsymbol{V}^{-1}\boldsymbol{B}^t(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{z}) \\
&\quad + 2(\boldsymbol{z}^t\boldsymbol{V}^{-1}\boldsymbol{B}^t(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{z} \\
&\quad - \boldsymbol{z}^t\boldsymbol{V}^{-1}\boldsymbol{B}^t(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\boldsymbol{F}_1(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1}\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{z})],
\end{aligned}
$$

where the expectation is taken with respect to $(\boldsymbol{B}, \boldsymbol{V}, \boldsymbol{z})$. *Here* $\boldsymbol{F}_1 = \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^t$ *and* $\boldsymbol{A} = -(q+1)\boldsymbol{I}_q + (\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1} + (\mathrm{tr}(\boldsymbol{I}_q + \boldsymbol{F}_1)^{-1})\boldsymbol{I}_q$.

Corollary 2.2 suggests that unbiased estimate of risk of the inverse regression estimator is the content of expectation of the right-hand side of (2.11) and, however, we seem unable to evaluate the risk difference of the classical and the inverse regression estimators analytically.

Table 1. Estimated risks (L_1) under multivariate normal distributions with $\boldsymbol{\xi} = (1,1,1,1,1)^t$.

$\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}^t$	CL	JS	Ave.	IN
$\mathbf{diag}(1,1,1,1,1)$	10.86	8.37	22.93%	4.77
	(0.241)	(0.156)		(0.006)
$\mathbf{diag}(10,10^{-1},10^{-1},10^{-1},10^{-1})$	8.34	7.45	10.71%	6.09
	(0.273)	(0.192)		(0.021)
$\mathbf{diag}(10,10,1,10^{-1},10^{-1})$	22.05	19.76	10.39%	12.66
	(0.720)	(0.616)		(0.036)
$\mathbf{diag}(10,10,10,10,10)$	51.18	47.56	7.07%	29.55
	(2.882)	(2.619)		(0.074)
$\mathbf{diag}(100^2,10,1,10^{-1},10^{-2})$	24.23	23.61	2.57%	13.11
	(0.877)	(0.840)		(0.064)
$\mathbf{diag}(10^2,10^2,10,10,1)$	44.58	43.80	1.74%	26.45
	(1.714)	(1.678)		(0.109)
$\mathbf{diag}(10^3,1,1,1,1)$	23.91	23.84	0.28%	7.09
	(1.408)	(1.399)		(0.045)
$\mathbf{diag}(10^3,10^2,10^2,10^2,10)$	35.00	34.89	0.32%	33.66
	(0.390)	(0.388)		(0.162)
$\mathbf{diag}(10^4,10^3,10^2,10,1)$	43.17	43.15	0.04%	23.98
	(1.061)	(1.060)		(0.131)
$\mathbf{diag}(10^4,10^4,10^3,10^2,10^2)$	31.90	31.89	0.02%	34.37
	(0.220)	(0.220)		(0.197)
$\mathbf{diag}(10^5,10^2,1,10^{-2},10^{-5})$	26.96	26.96	0.00%	11.25
	(0.796)	(0.796)		(0.085)

Remark 2.2. In Theorem 2.2, we have analytical dominance result on the classical and the shrinkage estimators when the quasi-loss function L is used. On the other hand we seem very difficult to establish the dominance result analytically under the loss function L_1 given in (1.10). Therefore, we have carried out Monte Carlo simulations to investigate risk performance under the loss L_1 which can be written as

$$L_1(\tilde{\boldsymbol{\xi}};\boldsymbol{\xi}) = (\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi})^t \boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}^t(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}). \tag{2.12}$$

The estimated risks are given in Tables 1 and 2 and our simulations are based on 10,000 independent replications which are generated from model (2.2). In Tables 1 and 2, 'CL', 'JS', and 'IN' denote the classical estimator (2.4), the James-Stein type estimator (2.10), and the inverse regression estimator (2.5), respectively, and their estimated standard deviations are in parentheses. Furthermore, 'Ave.' is average of improvement in risk of JS against that of CL, i.e., Ave. $= 100(1 - \widehat{R}^{*JS}/\widehat{R}^*)\%$, where $\widehat{R}^*$ and $\widehat{R}^{*JS}$ are, respectively, values of estimated risks for CL and JS by simulations. For simulations, we take $(n,m,p,q) = (30,20,7,5)$ and suppose that $\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}^t$ is the diagonal matrix with typical elements and that $\boldsymbol{\xi} = (1,1,1,1,1)^t$ (in Table 1) and $\boldsymbol{\xi} = (2,2,2,2,2)^t$ (in Table 2).

From numerical results in Tables 1 and 2, we observe that Ave.'s are large when the diagonal elements of $\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}^t$ are small. Hence, our simulations indicate that JS is better than CL under the loss L_1 but it is difficult to prove the domination analytically.

Table 2. Estimated risks (L_1) under multivariate normal distributions with $\boldsymbol{\xi} = (2,2,2,2,2)^t$.

$\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}^t$	CL	JS	Ave.	IN
$\mathbf{diag}(1,1,1,1,1)$	30.76	28.17	8.40%	18.89
	(0.628)	(0.502)		(0.018)
$\mathbf{diag}(10,10^{-1},10^{-1},10^{-1},10^{-1})$	22.17	21.68	2.20%	23.91
	(0.863)	(0.755)		(0.061)
$\mathbf{diag}(10,10,1,10^{-1},10^{-1})$	65.41	63.50	2.92%	49.66
	(2.052)	(1.941)		(0.115)
$\mathbf{diag}(10,10,10,10,10)$	169.99	166.60	1.99%	116.16
	(8.131)	(7.923)		(0.266)
$\mathbf{diag}(100^2,10,1,10^{-1},10^{-2})$	81.25	80.69	0.69%	49.51
	(9.692)	(9.576)		(0.200)
$\mathbf{diag}(10^2,10^2,10,10,1)$	151.29	150.60	0.46%	100.13
	(6.096)	(6.059)		(0.371)
$\mathbf{diag}(10^3,1,1,1,1)$	71.79	71.74	0.07%	24.57
	(5.679)	(5.669)		(0.125)
$\mathbf{diag}(10^3,10^2,10^2,10^2,10)$	122.54	122.43	0.08%	124.35
	(1.500)	(1.498)		(0.568)
$\mathbf{diag}(10^4,10^3,10^2,10,1)$	145.49	145.47	0.01%	86.36
	(3.566)	(3.566)		(0.446)
$\mathbf{diag}(10^4,10^4,10^3,10^2,10^2)$	111.76	111.75	0.01%	123.41
	(0.776)	(0.776)		(0.692)
$\mathbf{diag}(10^5,10^2,1,10^{-2},10^{-5})$	81.54	81.54	0.00%	38.66
	(2.421)	(2.421)		(0.268)

3. Extensions to elliptical errors

In this section we consider calibration problem under elliptical errors. Here, suppose that the error matrices, ϵ and ϵ_0, of (1.1) and (1.2) have a joint density function

$$|\boldsymbol{\Sigma}|^{-(n+m)/2} f(\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\epsilon}^t\boldsymbol{\epsilon} + \boldsymbol{\epsilon}_0^t\boldsymbol{\epsilon}_0)\}), \tag{3.1}$$

where f is an unknown function on $[0, \infty)$ and $\boldsymbol{\Sigma}$ is a $p \times p$ unknown positive-definite matrix. Note that the rows of both $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_0$ are uncorrelatedly distributed but not independently.

We shall state proofs of main theorems of this section in Subsection 4.2.

3.1 *A canonical form*

We first derive a canonical form of this setup. Let $\boldsymbol{\Upsilon}$ be an $n \times n$ orthogonal matrix such that

$$\boldsymbol{\Upsilon}\mathbf{1}_n = (n^{1/2}, 0, \ldots, 0)^t \quad \text{and} \quad \boldsymbol{\Upsilon}\boldsymbol{X} = [\mathbf{0}_{q\times 1}, (\boldsymbol{X}^t\boldsymbol{X})^{1/2}, \mathbf{0}_{q\times(n-q-1)}]^t.$$

Also let $\boldsymbol{\Upsilon}\boldsymbol{Y} = [n^{1/2}\boldsymbol{y}, \boldsymbol{B}^t, \boldsymbol{v}^t]^t$. Here the sizes of $\boldsymbol{y}$, $\boldsymbol{B}$ and $\boldsymbol{v}$ are, respectively, $p \times 1$, $q \times p$ and $(n - q - 1) \times p$. Similarly, let $\boldsymbol{\Upsilon}_0$ be an $m \times m$ orthogonal matrix such that $\boldsymbol{\Upsilon}_0\mathbf{1}_m = (m^{1/2}, 0, \ldots, 0)^t$ and denote $\boldsymbol{\Upsilon}_0\boldsymbol{Y}_0 = [m^{1/2}\boldsymbol{y}_0, \boldsymbol{v}_0^t]^t$, where the sizes of $\boldsymbol{y}_0$ and $\boldsymbol{v}_0$ are, respectively, $p \times 1$ and $(m - 1) \times p$. Thus, by the orthogonal transformations $\boldsymbol{Y} \to \boldsymbol{\Upsilon}\boldsymbol{Y}$ and $\boldsymbol{Y}_0 \to \boldsymbol{\Upsilon}_0\boldsymbol{Y}_0$, the density (3.1) can be written as

$$\begin{aligned}|\boldsymbol{\Sigma}|^{-(n+m)/2} f(\mathrm{tr}[\boldsymbol{\Sigma}^{-1}\{&n(\boldsymbol{y} - \boldsymbol{\alpha})(\boldsymbol{y} - \boldsymbol{\alpha})^t + (\boldsymbol{B} - \boldsymbol{\beta})^t(\boldsymbol{B} - \boldsymbol{\beta}) + \boldsymbol{v}^t\boldsymbol{v} \\ &+ m(\boldsymbol{y}_0 - \boldsymbol{\alpha} - c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})(\boldsymbol{y}_0 - \boldsymbol{\alpha} - c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})^t + \boldsymbol{v}_0^t\boldsymbol{v}_0\}]),\end{aligned} \tag{3.2}$$

where $\boldsymbol{\beta} = (\boldsymbol{X}^t\boldsymbol{X})^{1/2}\boldsymbol{\Theta}$, $\boldsymbol{\xi} = c_{n,m}^{-1/2}(\boldsymbol{X}^t\boldsymbol{X})^{-1/2}\boldsymbol{x}_0$, and $c_{n,m} = 1/n + 1/m$. Then our problem is to estimate $\boldsymbol{\xi}$ based on $(\boldsymbol{y}, \boldsymbol{B}, \boldsymbol{v}, \boldsymbol{v}_0, \boldsymbol{y}_0)$ with respect to the loss L given in (2.3).

3.2 *The classical estimator and its improved estimator*

Denote $\boldsymbol{S} = \boldsymbol{v}^t\boldsymbol{v} + \boldsymbol{v}_0^t\boldsymbol{v}_0$ and $\boldsymbol{z} = c_{n,m}^{-1/2}(\boldsymbol{y}_0 - \boldsymbol{y})$. If $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ are known and f is decreasing on $[0, \infty)$, then the maximum likelihood estimator of $\boldsymbol{\xi}$ is given by $\hat{\boldsymbol{\xi}} = c_{n,m}^{-1/2}(\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}^t)^{-1}\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_0 - \boldsymbol{\alpha})$. When $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ are unknown, we shall replace $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ by their estimators from the data $(\boldsymbol{y}, \boldsymbol{B}, \boldsymbol{S})$ without data $\boldsymbol{y}_0$. From (3.2) with a decreasing function f, the maximum likelihood estimator of $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ are $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}) = (\boldsymbol{y}, \boldsymbol{B}, \kappa\boldsymbol{S})$ where κ is a certain constant. Hence, we obtain a natural estimator

$$\begin{aligned}\hat{\boldsymbol{\xi}} &= c_{n,m}^{-1/2}(\hat{\boldsymbol{\beta}}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\beta}}^t)^{-1}\hat{\boldsymbol{\beta}}\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_0 - \hat{\boldsymbol{\alpha}}) \\ &= (\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}.\end{aligned} \tag{3.3}$$

Throughout this paper, this estimator is called the *classical estimator* in case of the elliptical model (3.2).

Consider an improvement on the classical estimator (3.3) with its extended estimators

$$\hat{\boldsymbol{\xi}}(\psi) = (1 - \psi/t)(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}, \tag{3.4}$$

where ψ is a differentiable function of $t = \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$. The estimators are extension of the estimators (2.9) in the case when the rows of errors $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_0$ follow the multivariate normal distributions.

Next, we shall evaluate risk of the estimators (3.4). Let g be a scalar-valued function of $(\boldsymbol{y}, \boldsymbol{B}, \boldsymbol{v}, \boldsymbol{v}_0, \boldsymbol{y}_0)$ and also let

$$F(t) = \frac{1}{2}\int_t^{+\infty} f(x)dx.$$

Denote

$$E_f[g] = \int g \times |\boldsymbol{\Sigma}|^{-(n+m)/2} f(\boldsymbol{x}) d\boldsymbol{y} d\boldsymbol{B} d\boldsymbol{v} d\boldsymbol{s}_0 d\boldsymbol{y}_0, \tag{3.5}$$

$$E_F[g] = \int g \times |\boldsymbol{\Sigma}|^{-(n+m)/2} F(\boldsymbol{x}) d\boldsymbol{y} d\boldsymbol{B} d\boldsymbol{v} d\boldsymbol{v}_0 d\boldsymbol{y}_0, \tag{3.6}$$

where $\boldsymbol{x} = \mathrm{tr}[\boldsymbol{\Sigma}^{-1}\{n(\boldsymbol{y}-\boldsymbol{\alpha})(\boldsymbol{y}-\boldsymbol{\alpha})^t + (\boldsymbol{B}-\boldsymbol{\beta})^t(\boldsymbol{B}-\boldsymbol{\beta}) + \boldsymbol{v}^t\boldsymbol{v} + m(\boldsymbol{y}_0 - \boldsymbol{\alpha} - c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})(\boldsymbol{y}_0 - \boldsymbol{\alpha} - c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})^t + \boldsymbol{v}_0^t\boldsymbol{v}_0\}]$. Using these notation, we give the risk expression of the estimators (3.4) as follows.

THEOREM 3.1. *Put* $t = \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$ *and* $l = n+m-q-2$. *Denote* $\psi' = d\psi/dt$. *Then, under the loss* L *given in* (2.3), *the risk of* $\hat{\boldsymbol{\xi}}(\psi)$ *can be written as*

$$\begin{aligned} R(\hat{\boldsymbol{\xi}}(\psi), \boldsymbol{\xi}) &= E_f[L(\hat{\boldsymbol{\xi}}(\psi); \boldsymbol{\xi})] \\ &= E_F[-p - 4(\psi'/t - \psi/t^2)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} + 2q(1-\psi/t) \\ &\quad + (l-p-1)(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - (1-\psi^2/t^2)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}) \\ &\quad - 4(\psi'/t - \psi/t^2)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} \\ &\quad \times (\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - (1-\psi/t)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}) \\ &\quad + 2q(1-\psi/t)(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})], \end{aligned}$$

provided a suitable condition is satisfied.

In Theorem 3.1, the content in $E_F[\cdot]$ is not unbiased estimate of risk in case of an elliptical density except normal density. The 'suitable condition' in Theorem 3.1 are the same as those of both Lemmas 4.7 and 4.8 in Subsection 4.2. From Theorem 3.1, we have an expression for risk of the classical estimator (3.3):

COROLLARY 3.1.

$$R(\hat{\boldsymbol{\xi}}, \boldsymbol{\xi}) = E_F[-p + 2q + (l-p-1+2q)(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})],$$

where $l = n+m-q-2$ *and* $\boldsymbol{F} = \boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t$.

Therefore, we get a dominance result under elliptical errors.

THEOREM 3.2. *Assume that we want to estimate* $\boldsymbol{\xi}$ *in* (3.2) *and that* $q \geq 3$. *If*
(i) ψ *is nondecreasing, and*

Table 3. Estimated risks (L_1) under multivariate t-distributions (joint) with $\boldsymbol{\xi} = (2,2,2,2,2)^t$.

$\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}^t$	CL	JS	Ave.	IN
$\mathbf{diag}(1,1,1,1,1)$	33.61	30.27	9.96%	18.92
	(2.706)	(1.923)		(0.019)
$\mathbf{diag}(10,10^{-1},10^{-1},10^{-1},10^{-1})$	24.83	24.18	2.64%	25.48
	(0.755)	(0.657)		(0.088)
$\mathbf{diag}(10,10,1,10^{-1},10^{-1})$	74.08	70.93	4.26%	52.75
	(4.821)	(4.421)		(0.173)
$\mathbf{diag}(10,10,10,10,10)$	208.59	201.76	3.28%	122.68
	(12.236)	(11.799)		(0.412)
$\mathbf{diag}(100^2,10,1,10^{-1},10^{-2})$	94.77	93.79	1.04%	70.39
	(3.626)	(3.533)		(0.546)
$\mathbf{diag}(10^2,10^2,10,10,1)$	215.29	212.71	1.20%	140.34
	(9.534)	(9.234)		(1.081)
$\mathbf{diag}(10^3,1,1,1,1)$	86.44	86.39	0.06%	37.25
	(6.428)	(6.400)		(0.589)
$\mathbf{diag}(10^3,10^2,10^2,10^2,10)$	215.26	214.72	0.25%	199.35
	(5.898)	(5.862)		(2.054)
$\mathbf{diag}(10^4,10^3,10^2,10,1)$	243.89	243.80	0.04%	130.14
	(15.338)	(15.327)		(1.431)
$\mathbf{diag}(10^4,10^4,10^3,10^2,10^2)$	206.49	206.42	0.04%	202.59
	(12.058)	(12.034)		(2.695)
$\mathbf{diag}(10^5,10^2,1,10^{-2},10^{-5})$	161.42	161.42	0.00%	63.42
	(36.580)	(36.578)		(0.780)

(ii) $0 \le \psi \le 2(q-2)/(l-p+3)$,
then the estimators (3.4) *improve on the classical estimator* (3.3) *under the loss* L.

The result of Theorem 3.2 is an extension of that of Theorem 2.2 and suggests that for our elliptically contoured distribution (3.1) we establish the robustness of improvement via the shrinkage estimators (2.9).

Although the risks of the estimators (2.5) and (2.6) can also be expressed by usage of notation (3.6), we omit these derivations.

3.3 *Monte Carlo studies*

Finally, using Monte Carlo simulations in special case of the parameters, we shall investigate the risk behavior of the improved estimators (3.4) under the loss function L_1 given in (2.12). We supposed that the errors are jointly distributed as a multivariate t-distribution whose density function is given by

$$(3.7) \qquad c_t|\boldsymbol{\Sigma}|^{-(n+m)/2}\{1 + (1/k)\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}^t\boldsymbol{\epsilon} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}_0^t\boldsymbol{\epsilon}_0)\}^{-(k+(n+m)p)/2},$$

where $c_t = \Gamma[\{k+(n+m)p\}/2]/\{(\pi k)^{(n+m)p/2}\Gamma[k/2]\}$ and $k > 0$. Here, we denote by $\Gamma(x)$ the Gamma function.

Our simulations are based on 10,000 independent replications which are generated from the canonical form (3.2). For this numerical studies we assume that $(n, m, p, q) =$

Table 4. Estimated risks (L_1) under multivariate t-distributions (i.i.d.) with $\boldsymbol{\xi} = (2,2,2,2,2)^t$.

$\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}^t$	CL	JS	Ave.	IN
$\mathbf{diag}(1,1,1,1,1)$	32.41	28.78	11.22%	19.19
	(0.874)	(0.696)		(0.017)
$\mathbf{diag}(10,10^{-1},10^{-1},10^{-1},10^{-1})$	26.10	25.33	2.94%	27.65
	(0.797)	(0.701)		(0.073)
$\mathbf{diag}(10,10,1,10^{-1},10^{-1})$	73.61	70.61	4.07%	57.06
	(2.579)	(2.344)		(0.136)
$\mathbf{diag}(10,10,10,10,10)$	231.04	224.48	2.84%	134.66
	(12.614)	(12.239)		(0.306)
$\mathbf{diag}(100^2,10,1,10^{-1},10^{-2})$	106.33	105.22	1.04%	69.51
	(9.048)	(8.863)		(0.353)
$\mathbf{diag}(10^2,10^2,10,10,1)$	219.90	218.45	0.66%	143.39
	(6.347)	(6.284)		(0.708)
$\mathbf{diag}(10^3,1,1,1,1)$	80.65	80.58	0.09%	31.97
	(2.428)	(2.423)		(0.247)
$\mathbf{diag}(10^3,10^2,10^2,10^2,10)$	219.38	219.11	0.12%	207.21
	(4.593)	(4.585)		(1.366)
$\mathbf{diag}(10^4,10^3,10^2,10,1)$	233.01	232.98	0.01%	136.22
	(6.432)	(6.431)		(1.051)
$\mathbf{diag}(10^4,10^4,10^3,10^2,10^2)$	192.30	192.28	0.01%	214.54
	(2.349)	(2.348)		(1.875)
$\mathbf{diag}(10^5,10^2,1,10^{-2},10^{-5})$	118.04	118.04	0.00%	61.50
	(7.568)	(7.568)		(0.502)

$(30,20,7,5)$ and that $k = 5$. We simulated the risks of the classical estimator (3.3), the James-Stein type shrinkage estimator with $\psi \equiv (q-2)/(l-p+3)$, and the inverse regression estimator $\check{\boldsymbol{\xi}} = (\boldsymbol{I}_q + \boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{V}^{-1}\boldsymbol{z}$ where $\boldsymbol{V} = \boldsymbol{v}^t\boldsymbol{v}$. These estimated risks are given in Table 3.

In Table 3, 'CL', 'JS', and 'IN' denote the classical, the James-Stein type, and the inverse regression estimators,. respectively, and their estimated standard deviations are in parentheses. Furthermore 'Ave.' indicates average of improvement in risk of JS against that of CL. We suppose that the parameter $\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}^t$ is the diagonal matrix with typical elements and that $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\xi} = (2,2,2,2,2)^t$.

Moreover, since the error distribution (3.7) does not denote independent sampling, we also conduct a simulation study based on independently and identically sampling model from the multivariate t-distribution. Here, its density function is given by

$$c_I|\Sigma|^{-1/2}\{1 + (1/k)\epsilon_i^t\Sigma^{-1}\epsilon_i\}^{-(k+p)/2}, \qquad i = 1,\ldots,n,n+1,\ldots,n+m,$$

where $c_I = \Gamma[(k+p)/2]/\{(\pi k)^{p/2}\Gamma[k/2]\}$, $\epsilon = [\epsilon_1,\ldots,\epsilon_n]^t$, and $\epsilon_0 = [\epsilon_{n+1},\ldots,\epsilon_{n+m}]^t$. For this simulation, the assumptions for (n,m,p,q,k) and parameter $(\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}^t)$ were the same as those in Table 3. This simulation result is given in Table 4.

From Table 3, we can see that JS performs better than CL in all cases and, particularly, Ave.'s are large when the diagonal elements of $\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}^t$ are small and close together. Thus, we seem that JS is better than CL even if the loss function L_1 is used. On the other hand the risk performance of Table 4 are similar to those in Table 3. Hence, although there are simulations in small cases of parameters, it is expected that the improvement with the estimator (3.4) remains robust under the loss L_1 even if all the rows of the error matrices ϵ and ϵ_0 are identically and independently distributed as an elliptically contoured distribution.

4. Proofs

4.1 *Proofs of Theorems* 2.1 *and* 2.2

First, to prove Theorem 2.1, we shall state definition of differential operators and calculation formulae with respect to the differential operators.

Let $\boldsymbol{z}$ be a $p\times 1$ vector and also let u and $\boldsymbol{u}$ be, respectively, scalar-valued and $p\times 1$ vector-valued functions of $\boldsymbol{z}$. Furthermore let $\boldsymbol{S}$ be a $p\times p$ symmetric, positive-definite matrix and let h and $\boldsymbol{H}$ be, respectively, scalar-valued and $p\times r$ matrix-valued functions of $\boldsymbol{S}$. Denote differential operators in terms of $\boldsymbol{z}=(z_i)$ and $\boldsymbol{S}=(S_{ij})$ by

$$\nabla_z; p\times 1=(\partial/\partial z_i) \quad \text{and} \quad D_S; p\times p=(\{D_S\}_{ij})=\left(\frac{1+\delta_{ij}}{2}\frac{\partial}{\partial S_{ij}}\right),$$

where δ_{ij} is Kronecker's delta. The actions of ∇_z on u and on $\boldsymbol{u}=(u_i)$ and those of D_S on h and on $\boldsymbol{H}=(H_{ij})$ are defined as, respectively,

$$\nabla_z u; p\times 1=\left(\frac{\partial u}{\partial z_i}\right), \quad \nabla_z \boldsymbol{u}^t; p\times p=\left(\frac{\partial u_j}{\partial z_i}\right), \quad \nabla_z^t \boldsymbol{u}; 1\times 1=\sum_{i=1}^{p}\frac{\partial u_i}{\partial z_i},$$

$$D_S h; p\times p=\left(\frac{1+\delta_{ij}}{2}\frac{\partial h}{\partial S_{ij}}\right), \quad \text{and} \quad D_S\boldsymbol{H}; p\times r=\left(\sum_{k=1}^{p}\frac{1+\delta_{ik}}{2}\frac{\partial H_{kj}}{\partial S_{ik}}\right).$$

Next we give the following lemmas in terms of calculus for operators ∇_z and D_S.

LEMMA 4.1. (Haff (1979, 1981, 1982)) *Let $\widetilde{D}_S$ be a $p\times p$ matrix whose elements are linear combinations of $\partial/\partial S_{ij}$ $(i=1,\dots,p, j=1,\dots,p)$. Also, let $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ be $p\times p$ matrices whose elements are functions of $\boldsymbol{S}$. Then we have*

(i) $\widetilde{D}_S\boldsymbol{H}_1\boldsymbol{H}_2=(\widetilde{D}_S\boldsymbol{H}_1)\boldsymbol{H}_2+(\boldsymbol{H}_1^t\widetilde{D}_S^t)^t\boldsymbol{H}_2$,
(ii) $D_S\boldsymbol{S}=\{(p+1)/2\}\boldsymbol{I}_p$,
(iii) $(\boldsymbol{H}_1 D_S)^t\boldsymbol{S}=\{\mathrm{tr}(\boldsymbol{H}_1)\}\boldsymbol{I}_p/2+\boldsymbol{H}_1/2$,
(iv) $\{D_S\}_{ij}S^{ab}=-(S^{aj}S^{ib}+S^{ai}S^{jb})/2$,

where S^{ab} is the (a,b)-element of $\boldsymbol{S}^{-1}$.

LEMMA 4.2. *Let ϕ be a function of $\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$ and also let $\boldsymbol{G}$ be a $q\times q$ matrix-valued function of $\boldsymbol{F}=\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t$, where $\boldsymbol{B}$ is a $q\times p$ matrix. Assume that $\boldsymbol{G}$ is symmetric. Furthermore, let D_F be a differential operator with respect to $\boldsymbol{F}$, i.e., $D_F; q\times q=(\{(1+\delta_{ij})/2\}\partial/\partial F_{ij})$. Then we have*

(i) $\mathrm{tr}[\nabla_z(\phi\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}-\boldsymbol{z})^t]=2\phi'\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}+\phi\,\mathrm{tr}(\boldsymbol{F}\boldsymbol{G})-p$,
(ii) $D_S\phi=-\phi'\boldsymbol{S}^{-1}\boldsymbol{z}\boldsymbol{z}^t\boldsymbol{S}^{-1}$,
(iii) $\{D_S\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\}_i=-\{\boldsymbol{S}^{-1}\boldsymbol{B}^t[(\boldsymbol{F}D_F)^t\boldsymbol{G}]\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\}_i-(1/2)\{\mathrm{tr}(\boldsymbol{F}\boldsymbol{G})\}\times\{\boldsymbol{S}^{-1}\boldsymbol{z}\}_i-(1/2)\{\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\}_i$,

where $\phi'=d\phi(t)/dt$ and $\{\boldsymbol{h}\}_i$ denotes the i-th element of a vector $\boldsymbol{h}$.

PROOF. (i) Now, it follows that $\nabla_z\phi = 2\phi' \boldsymbol{S}^{-1}\boldsymbol{z}$ and $\nabla_z \boldsymbol{z}^t = \boldsymbol{I}_p$. Thus we can see that

$$\begin{aligned}
&\mathrm{tr}[\nabla_z(\phi \boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} - \boldsymbol{z})^t] \\
&\quad = \mathrm{tr}[(\nabla_z\phi)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}] + \phi\,\mathrm{tr}[(\nabla_z\boldsymbol{z}^t)\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}] - \mathrm{tr}(\nabla_z\boldsymbol{z}^t) \\
&\quad = 2\phi'\,\mathrm{tr}(\boldsymbol{S}^{-1}\boldsymbol{z}\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}) + \phi\,\mathrm{tr}(\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}) - \mathrm{tr}(\boldsymbol{I}_p).
\end{aligned}$$

(ii) The (i,j)-element of $D_S\phi$ is equal to

$$\{D_S\}_{ij}\phi = \phi'\{D_S\}_{ij}(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}) = \phi'\sum_{a,b} z_a z_b \{D_S\}_{ij}S^{ab}.$$

Hence, from Lemma 4.1 (iv), we have the equality (ii).

(iii) It follows from Lemma 4.1 (i) that

$$\{D_S\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\}_i = \{(D_S\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B})\boldsymbol{S}^{-1}\boldsymbol{z}\}_i + \{(\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}D_S)^t\boldsymbol{S}^{-1}\boldsymbol{z}\}_i. \tag{4.1}$$

Applying Lemma 4.1 (iv) to the second term of the right-hand side in (4.1), we obtain

$$\begin{aligned}
&\{(\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}D_S)^t\boldsymbol{S}^{-1}\boldsymbol{z}\}_i \\
&\quad = \sum_{a,b,c}\{\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\}_{ab}(\{D_S\}_{ia}S^{bc})z_c \\
&\quad = -(1/2)[\mathrm{tr}(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G})]\{\boldsymbol{S}^{-1}\boldsymbol{z}\}_i - (1/2)\{\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\}_i.
\end{aligned} \tag{4.2}$$

Next, we evaluate the first term of the right-hand side in (4.1). We observe that

$$\begin{aligned}
\{(D_S\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B})\boldsymbol{S}^{-1}\boldsymbol{z}\}_i &= \sum_j [\{D_S\boldsymbol{B}^t\boldsymbol{G}\}_{ij}]\{\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\}_j \\
&= \sum_{j,a,b}[B_{ba}\{D_S\}_{ia}G_{bj}]\{\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\}_j.
\end{aligned} \tag{4.3}$$

Here, from chain rule and $\boldsymbol{F} = \boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t$, we get

$$\begin{aligned}
\sum_{a,b} B_{ba}\{D_S\}_{ia}G_{bj} &= \sum_{a,b,c,d} B_{ba}\cdot\left[\frac{1}{2}(1+\delta_{cd})\frac{\partial G_{bj}}{\partial F_{cd}}\right]\cdot\left[\frac{1}{2}(1+\delta_{ia})\frac{\partial F_{cd}}{\partial S_{ia}}\right] \\
&= \sum_{a,b,c,d,e,f} B_{ba}\cdot[\{D_F\}_{cd}G_{bj}]\cdot[B_{ce}B_{df}\{D_S\}_{ia}S^{ef}] \\
&= -\frac{1}{2}\sum_{a,b,c,d,e,f} B_{ba}\cdot[\{D_F\}_{cd}G_{bj}]\cdot B_{ce}B_{df}(S^{ea}S^{if} + S^{ei}S^{af}) \\
&= -\{\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^tD_F)^t\boldsymbol{G}\}_{ij},
\end{aligned}$$

where the third equality is given by Lemma 4.1 (iv). Hence, using the above result and (4.3), we can see that

$$\{(D_S\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B})\boldsymbol{S}^{-1}\boldsymbol{z}\}_i = -\{\boldsymbol{S}^{-1}\boldsymbol{B}^t\{(\boldsymbol{F}D_F)^t\boldsymbol{G}\}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\}_i. \tag{4.4}$$

Finally, combining (4.1), (4.2) and (4.4), we get the equality (iii). □

LEMMA 4.3. *Let ϕ, $\boldsymbol{F}$, and $\boldsymbol{G}$ be defined as in Lemma 4.2. Denote $\hat{\boldsymbol{\xi}}(\phi, \boldsymbol{G}) = \phi \boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}$. Then we have*

$$\begin{aligned}
&\operatorname{tr}[D_S(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})^t] \\
&\quad = 2\phi'\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - \phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}) \\
&\qquad + 2\phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{I}_q - \phi\boldsymbol{G}\boldsymbol{F})\{(\boldsymbol{F}D_{\boldsymbol{F}})^t\boldsymbol{G}\}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} \\
&\qquad + \phi[\operatorname{tr}(\boldsymbol{F}\boldsymbol{G})](\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - \phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}) \\
&\qquad + \phi(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} - \phi\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{F}\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}).
\end{aligned}$$

PROOF. We observe that

$$\begin{aligned}
&\operatorname{tr}[D_S(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})^t] \\
&\quad = 2\sum_i \{D_S(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})\}_i\{\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z}\}_i \\
&\quad = 2\sum_i \{(D_S\phi)\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} + \phi D_S\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\}_i\{\phi\boldsymbol{B}^t\boldsymbol{G}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} - \boldsymbol{z}\}_i.
\end{aligned} \tag{4.5}$$

Hence, in the first braces of the last right-hand side of (4.5), we apply Lemma 4.2 (ii) to the first term and Lemma 4.2 (iii) to the second term to obtain the desired results. □

Next, we state the Stein identity of the multivariate normal distribution and the Stein-Haff identity of the Wishart distribution for our problem. These identities are used to derive the unbiased estimate of risk for the estimators $\hat{\boldsymbol{\xi}}(\phi, \boldsymbol{G})$.

LEMMA 4.4. (Stein (1973)) *Let $\boldsymbol{z} \sim \mathcal{N}_p(\boldsymbol{\beta}^t\boldsymbol{\xi}, \boldsymbol{\Sigma})$. Also let $\boldsymbol{u}$ be a $p\times 1$ vector whose elements are differentiable functions of $\boldsymbol{z}$. Then we have*

$$E[(\boldsymbol{z} - \boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}] = E[\operatorname{tr}(\nabla_z\boldsymbol{u}^t)]$$

provided the expectations exist.

LEMMA 4.5. (Haff (1977)) *Let $\boldsymbol{S} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, l)$. Also let $\boldsymbol{H}$ be a $p \times p$ matrix whose elements are differentiable functions of $\boldsymbol{S}$. Then we have*

$$E[\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{H})] = E[(l-p-1)\operatorname{tr}(\boldsymbol{S}^{-1}\boldsymbol{H}) + 2\operatorname{tr}(D_S\boldsymbol{H})]$$

provided a suitable condition is satisfied.

PROOF OF THEOREM 2.1. From Lemmas 4.4 and 4.5, the risk of $\hat{\boldsymbol{\xi}}(\phi, \boldsymbol{G})$ under the loss L can be expressed as

$$\begin{aligned}
R(\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}),\boldsymbol{\xi}) &= E[L(\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}),\boldsymbol{\xi})] \\
&= E[(\boldsymbol{z} - \boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{z} - \boldsymbol{\beta}^t\boldsymbol{\xi}) + 2(\boldsymbol{z} - \boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z}) \\
&\qquad + \operatorname{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})^t\}] \\
&= E[p + 2\operatorname{tr}\{\nabla_z(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})^t\} \\
&\qquad + (l-p-1)\operatorname{tr}\{\boldsymbol{S}^{-1}(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})^t\} \\
&\qquad + 2\operatorname{tr}\{D_S(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\phi,\boldsymbol{G}) - \boldsymbol{z})^t\}].
\end{aligned}$$

Thus the desired result can be given by applying Lemma 4.2 (i) and Lemma 4.3, respectively, to the second and the last terms in the brackets of the last right-hand side of the above equality. □

Next, to evaluate risks of the classical and the inverse regression estimators, we give the following lemma.

LEMMA 4.6. *Let $\boldsymbol{F}$ be a $q \times q$ symmetric, positive-definite matrix. Then we have*
(i) $(\boldsymbol{F}D_F)^t\boldsymbol{F}^{-1} = -(q+1)\boldsymbol{F}^{-1}/2$,
(ii) $(\boldsymbol{F}D_F)^t(\boldsymbol{I}_q+\boldsymbol{F})^{-1} = -(1/2)(\boldsymbol{I}_q+\boldsymbol{F})^{-1}\{(q+1)\boldsymbol{I}_q - (\boldsymbol{I}_q+\boldsymbol{F})^{-1} - (\mathrm{tr}[(\boldsymbol{I}_q+\boldsymbol{F})^{-1}])\boldsymbol{I}_q\}$.

PROOF. (i) From Lemma 4.1 (i) and (ii), we can see that

$$\begin{aligned}\boldsymbol{0}_{q\times q} = D_F(\boldsymbol{F}\boldsymbol{F}^{-1}) &= (D_F\boldsymbol{F})\boldsymbol{F}^{-1} + (\boldsymbol{F}D_F)^t\boldsymbol{F}^{-1}\\ &= (q+1)\boldsymbol{F}^{-1}/2 + (\boldsymbol{F}D_F)^t\boldsymbol{F}^{-1}.\end{aligned}$$

Hence we have the equality (i).
(ii) Similarly, we observe that from Lemma 4.1 (i) and (ii)

$$\begin{aligned}\boldsymbol{0}_{q\times q} &= D_F\{(\boldsymbol{I}_q+\boldsymbol{F})(\boldsymbol{I}_q+\boldsymbol{F})^{-1}\}\\ &= (q+1)(\boldsymbol{I}_q+\boldsymbol{F})^{-1}/2 + \{(\boldsymbol{I}_q+\boldsymbol{F})D_F\}^t(\boldsymbol{I}_q+\boldsymbol{F})^{-1}\end{aligned}$$

and from Lemma 4.1 (i) and (iii)

$$\begin{aligned}\boldsymbol{0}_{q\times q} &= D_F\{(\boldsymbol{I}_q+\boldsymbol{F})^{-1}(\boldsymbol{I}_q+\boldsymbol{F})\}\\ &= \{D_F(\boldsymbol{I}_q+\boldsymbol{F})^{-1}\}(\boldsymbol{I}_q+\boldsymbol{F}) + (\mathrm{tr}[(\boldsymbol{I}_q+\boldsymbol{F})^{-1}])\boldsymbol{I}_q/2 + (\boldsymbol{I}_q+\boldsymbol{F})^{-1}/2.\end{aligned}$$

Thus, we can write the above equalities as, respectively,

$$\{(\boldsymbol{I}_q+\boldsymbol{F})D_F\}^t(\boldsymbol{I}_q+\boldsymbol{F})^{-1} = -(q+1)(\boldsymbol{I}_q+\boldsymbol{F})^{-1}/2, \tag{4.6}$$

$$D_F(\boldsymbol{I}_q+\boldsymbol{F})^{-1} = -(\mathrm{tr}[(\boldsymbol{I}_q+\boldsymbol{F})^{-1}])(\boldsymbol{I}_q+\boldsymbol{F})^{-1}/2 - (\boldsymbol{I}_q+\boldsymbol{F})^{-2}/2. \tag{4.7}$$

Here it follows that

$$\{(\boldsymbol{I}_q+\boldsymbol{F})D_F\}^t(\boldsymbol{I}_q+\boldsymbol{F})^{-1} = D_F(\boldsymbol{I}_q+\boldsymbol{F})^{-1} + (\boldsymbol{F}D_F)^t(\boldsymbol{I}_q+\boldsymbol{F})^{-1}. \tag{4.8}$$

Hence, combining (4.6)–(4.8), we obtain the equality (ii). □

PROOF OF COROLLARY 2.1. The proof is given easily from the combination of Theorem 2.1 and Lemma 4.6 (i). □

PROOF OF THEOREM 2.2. Under the loss L, the risk of the estimators (2.9) can be expressed as

$$\begin{aligned}R(\hat{\boldsymbol{\xi}}(\psi),\boldsymbol{\xi}) &= E[L(\hat{\boldsymbol{\xi}}(\psi);\boldsymbol{\xi})]\\ &= E[-p - 4(\psi'/t - \psi/t^2)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} + 2q(1-\psi/t)\\ &\quad + (l-p-1)(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - (1-\psi^2/t^2)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})\\ &\quad - 4(\psi'/t - \psi/t^2)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}\\ &\quad \times (\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - (1-\psi/t)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})\\ &\quad + 2q(1-\psi/t)(\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z} - \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z})],\end{aligned}$$

where $t = \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{z}$. Here, put $t_0 = \boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}$. Thus, the risk difference between $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}(\psi)$ can be written as

$$\begin{aligned}(4.9)\qquad & R(\hat{\boldsymbol{\xi}}(\psi),\boldsymbol{\xi}) - R(\hat{\boldsymbol{\xi}},\boldsymbol{\xi}) \\ &\quad = E[-4(\psi'/t - \psi/t^2)t_0 - 2q\psi/t + (l-p-1)\psi^2 t_0/t^2 \\ &\qquad - 4(\psi'/t - \psi/t^2)t_0\{t - (1-\psi/t)t_0\} - 2q\psi(t-t_0)/t].\end{aligned}$$

From the assumptions that $\psi \geq 0$ and that ψ is nondecreasing and the fact that $t - t_0 \geq 0$, the fourth term in brackets of the right-hand side of (4.9) can be evaluated as

$$\begin{aligned}(4.10)\qquad & -4(\psi'/t - \psi/t^2)t_0\{t - (1-\psi/t)t_0\} \\ &\quad = -4\psi' t_0(t - t_0 + \psi t_0/t)/t + 4\psi t_0(t-t_0)/t^2 + 4\psi^2 t_0^2/t^3 \\ &\quad \leq 0 + 4\psi(t-t_0)/t + 4\psi^2 t_0/t^2.\end{aligned}$$

Similarly, we obtain

$$(4.11)\qquad -4(\psi'/t - \psi/t^2)t_0 \leq 4\psi t_0/t^2 \quad \text{and} \quad -2q\psi/t \leq -2q\psi t_0/t^2.$$

Thus, combining (4.9)–(4.11), we have

$$\begin{aligned} R(\hat{\boldsymbol{\xi}}(\psi),\boldsymbol{\xi}) - R(\hat{\boldsymbol{\xi}},\boldsymbol{\xi}) &\leq E[4\psi t_0/t^2 - 2q\psi t_0/t^2 + (l-p-1)\psi^2 t_0/t^2 \\ &\qquad + 4\psi(t-t_0)/t + 4\psi^2 t_0/t^2 - 2q\psi(t-t_0)/t] \\ &= E[\{(l-p+3)\psi^2 - 2(q-2)\psi\}t_0/t^2 - 2(q-2)\psi(t-t_0)/t] \\ &\leq E[\{(l-p+3)\psi^2 - 2(q-2)\psi\}t_0/t^2].\end{aligned}$$

Hence, we complete the proof. □

Proof of Corollary 2.2. Replacing $\boldsymbol{S}$ by $\boldsymbol{V}$ in Theorem 2.1 and using Lemma 4.6 (ii), we can immediately get the desired result. □

4.2 *Proofs of Theorems* 3.1 *and* 3.2

In this subsection we give proofs of theorems and corollary in Section 3. The statistic $(\boldsymbol{y}, \boldsymbol{B}, \boldsymbol{v}, \boldsymbol{v}_0, \boldsymbol{y}_0)$ is the same defined as Section 3. First, we define the useful notation.

Let $\boldsymbol{u} = (u_1, \dots, u_p)^t$ be a $p \times 1$ vector whose elements are functions of $\boldsymbol{y} = (y_1, \dots, y_p)^t$ and $\boldsymbol{y}_0 = (y_{01}, \dots, y_{0p})^t$. Also let ∇_y and ∇_{y_0} be $p \times 1$ differential operators with respect to $\boldsymbol{y}$ and $\boldsymbol{y}_0$, respectively. Define

$$(4.12)\qquad (\nabla_y \boldsymbol{u}^t)_{ij} = \frac{\partial u_j}{\partial y_i} \quad \text{and} \quad (\nabla_{y_0}\boldsymbol{u}^t)_{ij} = \frac{\partial u_j}{\partial y_{0i}}$$

for $i = 1, \dots, p$, $j = 1, \dots, p$.

Further, let $\boldsymbol{W} \equiv \boldsymbol{W}(\boldsymbol{S}) = (W_{ij})$ be a $p \times p$ matrix such that the (i,j)-element W_{ij} is a function of $\boldsymbol{S} = (S_{ij})$. Let

$$(4.13)\qquad \{\boldsymbol{D}_S \boldsymbol{W}\}_{ij} = \sum_{a=1}^{p} d_{ia} W_{aj},$$

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 $i \neq a$. Put $\boldsymbol{v} = (\boldsymbol{v}_1^t, \dots, \boldsymbol{v}_{n-q-1}^t)^t$ and $\boldsymbol{v}_0 = (\boldsymbol{v}_{n-q}^t, \dots, \boldsymbol{v}_l^t)^t$ with $l = n+m-q-2$. Also put $\boldsymbol{v}_i = (v_{i1}, \dots, v_{ip})$ for $i = 1, \dots, l$. Hence we have $\boldsymbol{S} = \boldsymbol{v}^t\boldsymbol{v} + \boldsymbol{v}_0^t\boldsymbol{v}_0 = \sum_{i=1}^l \boldsymbol{v}_i^t\boldsymbol{v}_i$.

Now, we adapt the Stein and Stein-Haff identities with respect to the elliptically contoured distribution due to Kubokawa and Srivastava (1999, 2001) for our problem. Since the proofs are given in much similar way as in Kubokawa and Srivastava (1999, 2001), we state the following formulae without the proofs:

LEMMA 4.7. *Let $E_f[\cdot]$ and $E_F[\cdot]$ be defined as* (3.5) *and* (3.6), *respectively. Let $\boldsymbol{u}$ be a $p \times 1$ vector whose elements are functions of $(\boldsymbol{y}, \boldsymbol{y}_0)$. For $i = 1, \dots, p$, assume that the elements of $\boldsymbol{u}$ are differentiable with respect to y_i and y_{0i} and that*

(i) *$E_f[|(\boldsymbol{y}-\boldsymbol{\alpha})^t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}|]$ and $E_f[|(\boldsymbol{y}_0-\boldsymbol{\alpha}-c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}|]$ are finite;*

(ii) *$\lim_{y_i\to\pm\infty} \boldsymbol{u}\boldsymbol{y}^tF(y_i^2+a^2) = \boldsymbol{0}_{p\times p}$ and $\lim_{y_{0i}\to\pm\infty} \boldsymbol{u}\boldsymbol{y}_0^tF(y_{0i}^2+a^2) = \boldsymbol{0}_{p\times p}$ for any real a.*

Then we have

(i) $E_f[(\boldsymbol{y}-\boldsymbol{\alpha})^t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}] = E_F[\operatorname{tr}\{\nabla_y\boldsymbol{u}^t\}/n]$,

(ii) $E_f[(\boldsymbol{y}_0-\boldsymbol{\alpha}-c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}\boldsymbol{u}] = E_F[\operatorname{tr}\{\nabla_{y_0}\boldsymbol{u}^t\}/m]$.

LEMMA 4.8. *Let $\boldsymbol{W}$ be a $p\times p$ matrix whose elements are functions of $\boldsymbol{S} = \sum_{i=1}^l \boldsymbol{v}_i^t\boldsymbol{v}_i$. For $i = 1, \dots, l$, $j = 1, \dots, p$, assume that the elements of $\boldsymbol{W}$ are differentiable with respect to v_{ij} and that*

(i) *$E_f[|\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W})|]$ is finite;*

(ii) *$\lim_{v_{ij}\to\pm\infty} |v_{ij}|\boldsymbol{W}(\sum_{i=1}^l \boldsymbol{v}_i^t\boldsymbol{v}_i)^{-1}F(v_{ij}^2+a^2) = \boldsymbol{0}_{p\times p}$ for any real a.*

Then we have

$$E_f[\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W})] = E_F[(l-p-1)\operatorname{tr}(\boldsymbol{S}^{-1}\boldsymbol{W}) + 2\operatorname{tr}(\boldsymbol{D}_S\boldsymbol{W})].$$

Next, using Lemma 4.7, we get the following lemma to evaluate the risk of the estimators (3.4):

LEMMA 4.9. *Let $\hat{\boldsymbol{\xi}}(\psi)$ be defined as* (3.4). *Then we have*

(i) $E_f[(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})] = E_F[p]$,

(ii) $E_f[(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\psi)-\boldsymbol{z})^t] = E_F[-2(\psi'/t-\psi/t^2)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z} + (1-\psi/t)q-p]$

provided the conditions of Lemma 4.7 *are satisfied.*

PROOF. (i) From $\boldsymbol{z} = c_{n,m}^{-1/2}(\boldsymbol{y}_0-\boldsymbol{y})$, we observe that

$$\begin{aligned}(4.14)\qquad & E_f[(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})] \\ &= E_f[c_{n,m}^{-1}(\boldsymbol{y}_0-\boldsymbol{\alpha}-c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_0-\boldsymbol{\alpha}-c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi}) \\ &\quad - 2c_{n,m}^{-1}(\boldsymbol{y}_0-\boldsymbol{\alpha}-c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\alpha}) \\ &\quad + c_{n,m}^{-1}(\boldsymbol{y}-\boldsymbol{\alpha})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\alpha})].\end{aligned}$$

Hence, applying Lemma 4.7 to each term of the right-hand side in (4.14), we can see that

$$\begin{aligned}E_f[(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})] &= E_F[c_{n,m}^{-1}\operatorname{tr}\{\nabla_{y_0}(\boldsymbol{y}_0-\boldsymbol{\alpha}-c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})^t\}/m\\ &\qquad -2c_{n,m}^{-1}\operatorname{tr}\{\nabla_{y_0}(\boldsymbol{y}-\boldsymbol{\alpha})^t\}/m\\ &\qquad + c_{n,m}^{-1}\operatorname{tr}\{\nabla_{y}(\boldsymbol{y}-\boldsymbol{\alpha})^t\}/n]\\ &= E_F[p].\end{aligned}$$

(ii) First, we use Lemma 4.7 to get

$$\begin{aligned}&E_f[(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\psi)-\boldsymbol{z})]\\ &\quad= E_f[c_{n,m}^{-1}(\boldsymbol{y}_0-\boldsymbol{\alpha}-c_{n,m}^{1/2}\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}\\ &\qquad\times\{(1-\psi/t)\boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}-\boldsymbol{I}_p\}(\boldsymbol{y}_0-\boldsymbol{y})\\ &\qquad-c_{n,m}^{-1}(\boldsymbol{y}-\boldsymbol{\alpha})^t\boldsymbol{\Sigma}^{-1}\\ &\qquad\times\{(1-\psi/t)\boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}-\boldsymbol{I}_p\}(\boldsymbol{y}_0-\boldsymbol{y})]\\ &\quad= E_F[c_{n,m}^{-1}(1/m)\operatorname{tr}\{\nabla_{y_0}[(1-\psi/t)(\boldsymbol{y}_0-\boldsymbol{y})^t\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\\ &\qquad-(\boldsymbol{y}_0-\boldsymbol{y})^t]\}\\ &\qquad-c_{n,m}^{-1}(1/n)\operatorname{tr}\{\nabla_{y}[(1-\psi/t)(\boldsymbol{y}_0-\boldsymbol{y})^t\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\\ &\qquad-(\boldsymbol{y}_0-\boldsymbol{y})^t]\}].\end{aligned}\tag{4.15}$$

Here, the fact that $\nabla_{y_0}(\psi/t)=2c_{n,m}^{-1/2}(\psi'/t-\psi/t^2)\boldsymbol{S}^{-1}\boldsymbol{z}$ and $\nabla_y(\psi/t)=-2c_{n,m}^{-1/2}(\psi'/t-\psi/t^2)\boldsymbol{S}^{-1}\boldsymbol{z}$ yields

$$\begin{aligned}&\nabla_{y_0}[(1-\psi/t)(\boldsymbol{y}_0-\boldsymbol{y})^t\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}-(\boldsymbol{y}_0-\boldsymbol{y})^t]\\ &\quad=-\nabla_{y}[(1-\psi/t)(\boldsymbol{y}_0-\boldsymbol{y})^t\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}-(\boldsymbol{y}_0-\boldsymbol{y})^t]\\ &\quad\equiv-2(\psi'/t-\psi/t^2)\boldsymbol{S}^{-1}\boldsymbol{z}\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}\\ &\qquad+(1-\psi/t)\boldsymbol{S}^{-1}\boldsymbol{B}^t(\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{B}^t)^{-1}\boldsymbol{B}-\boldsymbol{I}_p.\end{aligned}\tag{4.16}$$

Hence, applying the above result (4.16) to the last right-hand side in (4.15), we have the equality (ii). □

PROOF OF THEOREM 3.1. From Lemmas 4.8 and 4.9, the risk of $\hat{\boldsymbol{\xi}}(\psi)$ under the loss L can be expressed as

$$\begin{aligned}R(\hat{\boldsymbol{\xi}}(\psi),\boldsymbol{\xi}) &= E_f[L(\hat{\boldsymbol{\xi}}(\psi),\boldsymbol{\xi})]\\ &= E_f[(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})+2(\boldsymbol{z}-\boldsymbol{\beta}^t\boldsymbol{\xi})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\psi)-\boldsymbol{z})\\ &\qquad+\operatorname{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\psi)-\boldsymbol{z})(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\psi)-\boldsymbol{z})^t\}]\\ &= E_F[-p-4(\psi'/t-\psi/t^2)\boldsymbol{z}^t\boldsymbol{S}^{-1}\boldsymbol{B}^t\boldsymbol{F}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{z}+2(1-\psi/t)q\\ &\qquad+(l-p-1)\operatorname{tr}\{\boldsymbol{S}^{-1}(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\psi)-\boldsymbol{z})(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\psi)-\boldsymbol{z})^t\}\\ &\qquad+2\operatorname{tr}\{\boldsymbol{D}_S(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\psi)-\boldsymbol{z})(\boldsymbol{B}^t\hat{\boldsymbol{\xi}}(\psi)-\boldsymbol{z})^t\}].\end{aligned}$$

Thus, applying Lemma 4.3 and Lemma 4.6 (i) to the last term in the brackets of the last right-hand side of the above equality, we get the risk expression of $\hat{\boldsymbol{\xi}}(\psi)$. □

PROOF OF COROLLARY 3.1. This is similar to proof of Corollary 2.1 and is omitted. □

PROOF OF THEOREM 3.2. This is similar to proof of Theorem 2.2 and is omitted. □

Acknowledgements

The author would like to thank Dr. Yoshihiko Konno for his helpful comments.

REFERENCES

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution, *Ann. Math. Statist.*, **41**, 642–645.

Branco, M., Bolfarine, H., Iglesias, P. and Arellano-Valle, R. B. (2000). Bayesian analysis of the calibration problem under elliptical distributions, *J. Statist. Plann. Inference*, **90**, 69–85.

Brown, P. J. (1982). Multivariate calibration (with discussion), *J. Roy. Statist. Soc. Ser. B*, **44**, 287–321.

Brown, P. J. (1993). *Measurement, Regression, and Calibration*, Oxford University Press, Oxford.

Haff, L. R. (1977). Minimax estimators for a multinormal precision matrix, *J. Multivariate Anal.*, **7**, 374–385.

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

Haff, L. R. (1981). Further identities for the Wishart distribution with applications in regression, *Canad. J. Statist.*, **9**, 215–224.

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

James, W. and Stein, C. (1961). Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 361–380, University of California Press, Berkeley.

Kubokawa, T. and Robert, C. P. (1994). New perspectives on linear calibration, *J. Multivariate Anal.*, **51**, 178–200.

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

Kubokawa, T. and Srivastava, M. S. (2001). Robust improvement in estimation of a mean matrix in an elliptically contoured distribution, *J. Multivariate Anal.*, **76**, 138–152.

Osborne, C. (1991). Statistical calibration: A review, *Internat. Statist. Rev.*, **59**, 309–336.

Rosenblatt, J. R. and Spiegelman, C. H. (1981). Discussion of Hunter and Lamboy's paper, *Technometrics*, **23**, 329–333.

Srivastava, M. S. (1995). Comparison of the inverse and classical estimators in multi-univariate linear calibration, *Comm. Statist. Theory Methods*, **24**, 2753–2767.

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

Sundberg, R. (1999). Multivariate calibration—Direct and indirect regression methodology (with discussion), *Scand. J. Statist.*, **26**, 161–207.

Tsukuma, H. (2002). A note on estimation under the quadratic loss in multivariate calibration, *J. Japan Statist. Soc.*, **32**, 165–181.