The Stein Phenomenon for Monotone Incomplete Multivariate Normal Data

Donald St. P. Richards* and Tomoya Yamada†

March 3, 2008

Abstract

We establish the Stein phenomenon in the context of two-step, monotone incomplete data drawn from $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a multivariate normal population with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. On the basis of data consisting of n observations on all $p+q$ characteristics and an additional $N-n$ observations on the last q characteristics, where all observations are mutually independent, denote by $\widehat{\boldsymbol{\mu}}$ the maximum likelihood estimator of $\boldsymbol{\mu}$. We establish criteria which imply that shrinkage estimators of James-Stein type have lower risk than $\widehat{\boldsymbol{\mu}}$ under Euclidean quadratic loss. Further, we show that the corresponding positive-part estimators improve on their unrestricted counterparts. We derive results for the case in which $\boldsymbol{\Sigma}$ is block-diagonal, the loss function is quadratic and non-spherical, and the shrinkage estimator is constructed by means of a non-decreasing, differentiable function of a quadratic form in $\widehat{\boldsymbol{\mu}}$. In the case of the problem of shrinking $\widehat{\boldsymbol{\mu}}$ to a vector whose components have a common value constructed from the data, we derive improved shrinkage estimators and again determine conditions under which the positive-part analogs have lower risk than their unrestricted counterparts.

1 Introduction

The Stein phenomenon, its appearance over sixty years ago notwithstanding, remains today a remarkable result: Given a random sample from a d-dimensional multivariate normal population, the sample mean, which is the "natural" and the maximum likelihood estimator of the population mean, is inadmissible with respect to quadratic loss for $d \geq 3$. In the ensuing decades since this revolutionary result was brought to light by Stein [32] and James and Stein [24], the phenomenon has engendered a literature of enormous size and scope, so that Stein's shrinkage estimators and their descendants are utilized today in many aspects of statistical theory and applications. Consequently, the Stein phenomenon exhibits a certain universality in nature, in the sense that it occurs for many loss functions, many inference problems, and many distributions. For bibliographies on the field, we refer to the lecture notes of Brown [14] and the monographs of Arnold [4], Berger [8], Brown [12], Casella and Berger [15], and Judge and Bock [26], these being only a few of the many books on the subject.

In this paper, we study the Stein phenomenon for $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a multivariate normal population with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We shall derive improved estimators

---

*Department of Statistics, Penn State University, University Park, PA 16802, U.S.A.

*Supported in part by National Science Foundation grant DMS-0705210.

†Department of Economics, Sapporo Gakuin University, 11 Bunkyodai, Ebetsu, Hokkaido, Japan.

†Research supported by a sabbatical leave-of-absence from Sapporo Gakuin University.

2000 Mathematics Subject Classification: Primary 62C15, 62H10; Secondary 60D10, 62E15.

Key words and phrases. Cauchy's interlacing theorem; empirical Bayes estimation; James-Stein estimator; missing data; positive-part estimator; shrinkage estimator; squared-error loss function; Wishart distribution.

for $\boldsymbol{\mu}$ when the data are *two-step monotone incomplete*, consisting of N mutually independent vectors, n of which are complete observations on all $p+q$ population characteristics and the remaining $N-n$ data vectors are on the last q characteristics only. We write this data in the form

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{Y}_1 \end{pmatrix} \ \begin{pmatrix} \boldsymbol{X}_2 \\ \boldsymbol{Y}_2 \end{pmatrix} \ \cdots \ \begin{pmatrix} \boldsymbol{X}_n \\ \boldsymbol{Y}_n \end{pmatrix} \ \boldsymbol{Y}_{n+1} \ \ \boldsymbol{Y}_{n+2} \ \ \cdots \ \ \boldsymbol{Y}_N \tag{1.1}$$

where each $\boldsymbol{X}_j$ is a $p \times 1$ vector and each $\boldsymbol{Y}_j$ is $q \times 1$. Monotone incomplete multivariate normal samples have been widely studied; see [1, 6, 7, 9, 16, 17, 18, 22, 27, 29, 30] and the many references provided in those papers.

On being given data of the form (1.1) it is well-known that $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, may be obtained by means of explicit, closed-form expressions (see Wu and Perlman [35], and references therein, for general *lattice conditional models* wherein general monotone data structures lead to explicit formulas for maximum likelihood estimators). The estimator $\widehat{\boldsymbol{\mu}}$ has many pleasant features, including unbiasedness, and those features also raise the issue of whether or not the Stein phenomenon holds for $\widehat{\boldsymbol{\mu}}$. Thus, if the quadratic loss function,

$$L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) = \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|^2, \tag{1.2}$$

measures the error resulting from estimating a parameter $\boldsymbol{\theta} \in \mathbb{R}^p$ by a statistic $\widehat{\boldsymbol{\theta}}$, and if $R(\boldsymbol{\theta}; \widehat{\boldsymbol{\theta}}) = E\big(L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}})\big)$ is the corresponding *risk function*, then we wish to find a shrinkage estimator, $\boldsymbol{m}(\widehat{\boldsymbol{\mu}})$ such that $R(\boldsymbol{\mu}; \widehat{\boldsymbol{\mu}}) > R\big(\boldsymbol{\mu}; \boldsymbol{m}(\widehat{\boldsymbol{\mu}})\big)$ for all $\boldsymbol{\mu} \in \mathbb{R}^p$.

Let $\boldsymbol{\mu}_{\boldsymbol{Y}}$ denote the mean of $\boldsymbol{Y}_1$ in (1.1); then it is well-known that, on the basis of the monotone sample (1.1), the maximum likelihood estimator of $\boldsymbol{\mu}_{\boldsymbol{Y}}$ is $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} = \bar{\boldsymbol{Y}} \equiv N^{-1}(\boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_N)$. Using only a simple property of $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$, we can obtain improved estimation for $\boldsymbol{\mu}$ under squared-error loss if the shrinkage target vector $\boldsymbol{\nu} \in \mathbb{R}^{p+q}$ is sufficiently close to $\boldsymbol{\mu}$. Specifically, define an estimator of James-Stein type,

$$\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c) = \left(1 - \frac{c}{\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^2}\right)(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}) + \boldsymbol{\nu}, \tag{1.3}$$

where $c > 0$ is a constant. For $q \geq 3$ and $\operatorname{Cov}(\boldsymbol{Y}_1) = \boldsymbol{I}_q$, we shall use the elementary result that $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$ is normally distributed to show at the outset of Section 3 that if $\|\boldsymbol{\mu} - \boldsymbol{\nu}\|^2 < (q-2)/N$ and

$$c^* = \frac{q-2}{N}\left(1 - \Big(\frac{N}{q-2}\Big)^{1/2} \|\boldsymbol{\mu} - \boldsymbol{\nu}\|\right) \tag{1.4}$$

then

$$R(\boldsymbol{\mu}; \widehat{\boldsymbol{\mu}}) - R\big(\boldsymbol{\mu}; \boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)\big) > 0, \tag{1.5}$$

for all $c \in (0, 2c^*)$; furthermore, this difference is maximized at $c = c^*$. The inequality (1.5) makes plausible the possibility that $\widehat{\boldsymbol{\mu}}$ can be improved uniformly by an estimator of James-Stein type and therefore is inadmissible.

To extend (1.5) to the case in which $\boldsymbol{\nu}$ is arbitrarily distant from $\boldsymbol{\mu}$ requires a more delicate analysis involving the exact distribution of $\widehat{\boldsymbol{\mu}}$. Indeed, the fundamental reason for the hitherto unavailability of shrinkage estimators of $\boldsymbol{\mu}$ with monotone incomplete data was the lack of knowledge of the exact distribution of $\widehat{\boldsymbol{\mu}}$; because of that impediment, it was not possible to derive explicit analytical expressions or bounds for $R\big(\boldsymbol{\mu}; \boldsymbol{m}(\widehat{\boldsymbol{\mu}})\big)$, the risk of the shrinkage estimator in (1.3). Recently, the distribution of $\widehat{\boldsymbol{\mu}}$ has been derived by Chang and Richards [16], and it is the purpose of this paper to exploit that result by constructing improved estimators of $\widehat{\boldsymbol{\mu}}$, proving that the Stein phenomenon prevails in the monotone incomplete setting. Remarkably, the estimator (1.3) plays a central role in the results to follow, and this leads us to speculate in Section 6 that, there may be an aspect of universality even to the estimator (1.3) within the context of arbitrarily-patterned incomplete multivariate normal samples.

In reviewing the literature on the Stein phenomenon, we have found results on shrinkage estimators of the population covariance matrix with incomplete data; see, e.g., Konno [27], and Sharma and Krishnamoorthy [31]. On the other hand, there seems to have been a nearly complete absence of results on shrinkage estimation for $\boldsymbol{\mu}$ in the same context. To the best of our knowledge, the only result on such shrinkage estimation for $\boldsymbol{\mu}$ is contained in the commentary of Fienberg [20] in his discussion of the paper by Hartley and Hocking [23]. In the case of a two-step monotone incomplete multivariate normal sample, Fienberg noted the normality of $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$, applied the James-Stein method to derive an estimator having lower risk than $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$, and posed the problem of improving $\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}$, the maximum likelihood estimator of the first p components of $\boldsymbol{\mu}$. Since that time, the problem had remained unaddressed.

Before closing the introduction, we emphasize that although the results of this paper are in line with those which may have been expected from a reading of the classical literature on shrinkage estimation, the derivation of these results are not straightforward extensions of classical arguments. Indeed, to establish certain shrinkage phenomena requires results from matrix analysis perhaps more intricate than those commonly arising in the area of shrinkage estimation, *e.g.*, Cauchy's Theorem on the interlacing properties of the eigenvalues of a principal submatrix of a positive definite matrix. In our view, the fact that the Stein phenomenon prevails within the context of monotone incomplete multivariate normal data reinforces the apparent universality of the phenomenon.

We close here with a description of the organization of the ensuing results. In Section 2, we provide some details and results necessary for the subsequent development. In Section 3, we consider the case in which $\boldsymbol{\Sigma}$ is a scalar matrix, extending to monotone incomplete samples the famous result of James and Stein [24] on the inadmissibility of the sample mean and also the later extension by Baranchik [5] the improvement involving positive-part estimators of James-Stein type. In Section 4, we consider the case in which $\boldsymbol{\Sigma}$ is diagonal or block-diagonal, and the shrinkage estimator may possibly depend on non-Euclidean distances between $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{\nu}$. In Section 5, we extend results of Lindley and Smith [28] and Efron and Morris [19] in which the shrinkage target $\boldsymbol{\nu}$ is constructed from the data. Finally, in Section 6, we comment upon some open problems raised by our results.

2 Preliminaries

Throughout the paper, matrices and vectors are represented by boldface type. In particular, we denote the identity matrix of order d by $\boldsymbol{I}_d$, and we also denote by $\mathbf{0}$ any matrix or vector of zeros, the dimension of which will be clear from the context.

We partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in conformity with (1.1), writing

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{X}} \\ \boldsymbol{\mu}_{\boldsymbol{Y}} \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\boldsymbol{\mu}_{\boldsymbol{X}} = (\mu_1, \ldots, \mu_p)'$, $\boldsymbol{\mu}_{\boldsymbol{Y}} = (\mu_{p+1}, \ldots, \mu_{p+q})'$; and $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, and $\boldsymbol{\Sigma}_{22}$ are of order $p \times p$, $p \times q$, and $q \times q$, respectively. We assume throughout that $n \geq q + 3$ to ensure that all expectations encountered later are absolutely convergent. We also use the notation $\tau = n/N$

for the proportion of data which are complete; and we denote $1-\tau$ by $\bar{\tau}$, so that $\bar{\tau}=(N-n)/N$ is the proportion of incomplete observations.

Define sample means

$$\bar{\boldsymbol{X}}=\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{X}_j,\quad \bar{\boldsymbol{Y}}_1=\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{Y}_j,\quad \bar{\boldsymbol{Y}}_2=\frac{1}{N-n}\sum_{j=n+1}^{N}\boldsymbol{Y}_j,\quad \bar{\boldsymbol{Y}}=\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{Y}_j, \tag{2.1}$$

and the corresponding matrices of sums of squares and products by

$$\begin{aligned}\boldsymbol{A}_{11}&=\sum_{j=1}^{n}(\boldsymbol{X}_j-\bar{\boldsymbol{X}})(\boldsymbol{X}_j-\bar{\boldsymbol{X}})', & \boldsymbol{A}_{12}=\boldsymbol{A}_{21}'&=\sum_{j=1}^{n}(\boldsymbol{X}_j-\bar{\boldsymbol{X}})(\boldsymbol{Y}_j-\bar{\boldsymbol{Y}}_1)',\\ \boldsymbol{A}_{22,n}&=\sum_{j=1}^{n}(\boldsymbol{Y}_j-\bar{\boldsymbol{Y}}_1)(\boldsymbol{Y}_j-\bar{\boldsymbol{Y}}_1)', & \boldsymbol{A}_{22,N}&=\sum_{j=1}^{N}(\boldsymbol{Y}_j-\bar{\boldsymbol{Y}})(\boldsymbol{Y}_j-\bar{\boldsymbol{Y}})'.\end{aligned} \tag{2.2}$$

It is well-known (*cf.* [1, 3, 30, 25]) that the maximum likelihood estimator of $\boldsymbol{\mu}$ is $\widehat{\boldsymbol{\mu}}=\begin{pmatrix}\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}\\ \widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}\end{pmatrix}$ where

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}=\bar{\boldsymbol{X}}-\bar{\tau}\boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}(\bar{\boldsymbol{Y}}_1-\bar{\boldsymbol{Y}}_2),\qquad \widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}=\bar{\boldsymbol{Y}}. \tag{2.3}$$

Let

$$\boldsymbol{\Omega}=\frac{1}{N}\boldsymbol{\Sigma}+\frac{\bar{\tau}}{n}\begin{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} \tag{2.4}$$

where $\boldsymbol{\Sigma}_{11\cdot 2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. Then we have the following result.

Theorem 2.1. (Chang and Richards [16]) *Let $\boldsymbol{V}_1\sim N_{p+q}(\mathbf{0},\boldsymbol{\Omega})$, $Q_1\sim\chi^2_{n-q}$, $Q_2\sim\chi^2_q$, $\boldsymbol{V}_2\sim N_p(\mathbf{0},\boldsymbol{I}_p)$, where $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, Q_1, and Q_2 are mutually independent. Then $\widehat{\boldsymbol{\mu}}$ satisfies the stochastic representation,*

$$\widehat{\boldsymbol{\mu}}\overset{\mathcal{L}}{=}\boldsymbol{\mu}+\boldsymbol{V}_1+\sqrt{\frac{\bar{\tau}Q}{n}}\begin{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{V}_2\\ \mathbf{0}\end{pmatrix}, \tag{2.5}$$

where $Q=Q_2/Q_1$. In particular, $\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}$ and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$ are mutually independent if and only if $\boldsymbol{\Sigma}_{12}=\mathbf{0}$.

Noting that $E(Q)=E(Q_2)\,E(1/Q_1)=q/(n-q-2)$, it follows from (2.5) that

$$\boldsymbol{\Sigma}_*\equiv\operatorname{Cov}(\widehat{\boldsymbol{\mu}})=\boldsymbol{\Omega}+\frac{q\bar{\tau}}{n(n-q-2)}\begin{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{V}_2\\ \mathbf{0}\end{pmatrix}. \tag{2.6}$$

For ease of future reference, we state explicitly a consequence of (2.5) that will be utilized repeatedly in the sequel.

Corollary 2.2. *Conditional on $Q=Q_2/Q_1$, the distribution of $\widehat{\boldsymbol{\mu}}$ is*

$$N_{p+q}\left(\boldsymbol{\mu},\boldsymbol{\Omega}+\frac{\bar{\tau}Q}{n}\begin{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\right). \tag{2.7}$$

In particular, if $\boldsymbol{\Sigma}=\boldsymbol{I}_{p+q}$ then $\widehat{\boldsymbol{\mu}}|Q\sim N_{p+q}(\boldsymbol{\mu},\boldsymbol{C}_Q)$ where

$$\boldsymbol{C}_Q=\begin{pmatrix}\frac{1}{n}(1+\bar{\tau}Q)\boldsymbol{I}_p & \mathbf{0}\\ \mathbf{0} & \frac{1}{N}\boldsymbol{I}_q\end{pmatrix}. \tag{2.8}$$

We shall also need Stein's fundamental integration-by-parts lemma. In stating that result, for a continuously differentiable function $\boldsymbol{\Psi} : \mathbb{R}^d \to \mathbb{R}^d$ and $\boldsymbol{z} = (z_1, \ldots, z_d)' \in \mathbb{R}^d$, we use the notation

$$\nabla \cdot \boldsymbol{\Psi}(\boldsymbol{z}) = \sum_{j=1}^{d} \frac{\partial}{\partial z_j} \boldsymbol{\Psi}(\boldsymbol{z}).$$

Lemma 2.3. (Stein [33, 34]) *Suppose that* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)' \in \mathbb{R}^d$, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ *is a diagonal positive definite matrix,* $\psi : \mathbb{R}^d \to \mathbb{R}$ *is absolutely continuous, and* $\psi_j(\boldsymbol{z}) \equiv \partial\psi(\boldsymbol{z})/\partial z_j$, $j = 1, \ldots, d$. *If* $\boldsymbol{Z} = (Z_1, \ldots, Z_d)' \sim N_d(\boldsymbol{\theta}, \boldsymbol{\Lambda})$ *then*

$$E\big((Z_j - \theta_j)\psi(\boldsymbol{Z})\big) = \lambda_j \, E\, \psi_j(\boldsymbol{Z}), \tag{2.9}$$

$j = 1, \ldots, d$, *as long as the expectations exist. In particular, if* $\boldsymbol{\Lambda} = \lambda^2 \boldsymbol{I}_d$ *and each coordinate function of* $\boldsymbol{\Psi} : \mathbb{R}^d \to \mathbb{R}^d$ *is absolutely continuous then*

$$E\big((\boldsymbol{Z} - \boldsymbol{\theta})'\boldsymbol{\Psi}(\boldsymbol{Z})\big) = \lambda^2 \, E\big(\nabla \cdot \boldsymbol{\Psi}(\boldsymbol{Z})\big) \tag{2.10}$$

We shall need a preliminary result on the expected value of the inverse of a multivariate normal quadratic form.

Lemma 2.4. *Let* $\boldsymbol{V} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ *where* $\boldsymbol{\Lambda}$ *is nonsingular, and denote by* $\lambda_{(d)}$ *the smallest eigenvalue of* $\boldsymbol{\Lambda}$. *Then, for* $d \geq 3$, $E(\boldsymbol{V}'\boldsymbol{V})^{-1} \leq \big((d-2)\lambda_{(d)}\big)^{-1}$.

Proof. Let $\boldsymbol{H}$ be a $d \times d$ orthogonal matrix such that $\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ is diagonal, say, $\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}' = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$. Making the transformation from $\boldsymbol{V}$ to $\boldsymbol{H}\boldsymbol{V}$, and noting that the quadratic form $\boldsymbol{V}'\boldsymbol{V}$ is invariant under this transformation, we shall assume, with no loss of generality that Λ is diagonal. Then $V_1, \ldots, V_d$, the components of $\boldsymbol{V}$ are mutually independent with $V_j \sim N(\mu_j, \lambda_j)$, $j = 1, \ldots, d$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)'$, and therefore $V_j^2/\lambda_j \sim \chi_1^2(\mu_j^2)$, a noncentral chi-square distribution with one degree-of-freedom and noncentrality parameter μ_j^2. Since $\lambda_j \geq \lambda_{(d)}$, $j = 1, \ldots, d$ then

$$\boldsymbol{V}'\boldsymbol{V} \overset{\mathcal{L}}{=} \sum_{j=1}^{d} \lambda_j \chi_1^2(\mu_j^2) \geq \lambda_{(d)} \sum_{j=1}^{d} \chi_1^2(\mu_j^2) \overset{\mathcal{L}}{=} \lambda_{(d)} \chi_d^2(\mu_1^2 + \cdots + \mu_d^2),$$

where the last equality holds since the noncentral chi-square distribution is additive in its degrees-of-freedom and in its noncentrality parameter. Therefore $\boldsymbol{V}'\boldsymbol{V} \overset{\mathcal{L}}{\geq} \lambda_{(d)}\chi_d^2(\boldsymbol{\mu}'\boldsymbol{\mu})$, and we now obtain

$$E(\boldsymbol{V}'\boldsymbol{V})^{-1} \leq \lambda_{(d)}^{-1} \, E\left(1/\chi_d^2(\boldsymbol{\mu}'\boldsymbol{\mu})\right).$$

By Anderson [2], p. 95, eq. (11),

$$E\left(1/\chi_d^2(\boldsymbol{\mu}'\boldsymbol{\mu})\right) = e^{-\boldsymbol{\mu}'\boldsymbol{\mu}/2} \sum_{j=0}^{\infty} \frac{(\boldsymbol{\mu}'\boldsymbol{\mu}/2)^j}{j!} \, \frac{1}{d+2j-2}, \tag{2.11}$$

$d \geq 3$. Since $1/(d+2j-2) \leq 1/(d-2)$ for $j \geq 0$ then the right-hand side of (2.11) is bounded above by

$$\frac{1}{d-2} \, e^{-\boldsymbol{\mu}'\boldsymbol{\mu}/2} \sum_{j=0}^{\infty} \frac{(\boldsymbol{\mu}'\boldsymbol{\mu}/2)^j}{j!} = \frac{1}{d-2},$$

and the proof now is complete. $\square$

Before turning to the statements and proofs of the main results, we comment on the basic strategy underlying the details of those proofs. As a consequence of the stochastic representation (2.5) we deduce, first, that $\widehat{\boldsymbol{\mu}}$, conditional on Q, has a multivariate normal distribution. Second, we shall apply Stein's Lemma to the conditional distribution of $\widehat{\boldsymbol{\mu}}$ given Q to obtain a lower bound on the difference in risks conditional on Q, and then we shall calculate the expectation with respect to Q to derive a lower bound for the overall difference in risks. We remark that although this strategy can be described in a straightforward manner, the technical details required to extend various classical results appear to be nontrivial.

3 The case of scalar covariance matrices

We begin by establishing (1.5). Define

$$\begin{aligned}\Delta R(\boldsymbol{\mu}) &= E\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\|^2 - E\|\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c)-\boldsymbol{\mu}\|^2 \\ &= E\left\{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\|^2 - \left\|(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}) - \frac{c}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu})\right\|^2\right\},\end{aligned} \tag{3.1}$$

the incremental risk in $\widehat{\boldsymbol{\mu}}$ over $\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c)$. On applying the parallelogram law,

$$\|\boldsymbol{u}\|^2 - \|\boldsymbol{u}-\boldsymbol{v}\|^2 = 2\boldsymbol{u}'\boldsymbol{v} - \|\boldsymbol{v}\|^2, \tag{3.2}$$

$\boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^d$, to expand (3.1) and simplifying the resulting expression, we obtain

$$\Delta R(\boldsymbol{\mu}) = E\left\{\frac{2c}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu})'(\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}) - \frac{c^2}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}\right\}. \tag{3.3}$$

Let $\boldsymbol{\kappa} = \boldsymbol{\mu}-\boldsymbol{\nu}$; then

$$\begin{aligned}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu})'(\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}) &= (\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}-\boldsymbol{\kappa})'(\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}) \\ &= \|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2 - \boldsymbol{\kappa}'(\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}) \\ &\ge \|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2 - \|\boldsymbol{\kappa}\|\cdot\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|,\end{aligned}$$

where the inequality follows from the Cauchy-Schwartz inequality. Therefore,

$$\Delta R(\boldsymbol{\mu}) \ge E\left\{2c\left(1-\frac{\|\boldsymbol{\kappa}\|}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|}\right) - \frac{c^2}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}\right\}. \tag{3.4}$$

Let $\boldsymbol{\nu}_2$ be the vector containing the last q components of $\boldsymbol{\nu}$. It is not difficult to see that $\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2 \ge \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2\|^2$; therefore $\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^{-j} \le \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2\|^{-j}$, $j=1,2$. Since $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} \sim N_q(\boldsymbol{\mu}_{\boldsymbol{Y}}, \frac{1}{N}\boldsymbol{I}_q)$ then $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2 \sim N_q(\boldsymbol{\mu}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2, \frac{1}{N}\boldsymbol{I}_q)$. Therefore, by Lemma 2.4,

$$E\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2\|^{-2} = E\big((\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2)'(\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2)\big)^{-1} < \frac{N}{q-2}, \tag{3.5}$$

and in turn, by the Cauchy-Schwartz inequality,

$$E\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^{-1} \le \big(E\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^{-2}\big)^{1/2} < \left(\frac{N}{q-2}\right)^{1/2}. \tag{3.6}$$

Substituting (3.5) and (3.6) into (3.4), we have

$$\Delta R(\boldsymbol{\mu}) > c\left\{2\left(1-\left(\frac{N}{q-2}\right)^{1/2}\|\boldsymbol{\kappa}\|\right)-\frac{N}{q-2}c\right\}. \tag{3.7}$$

It now follows that if $\|\boldsymbol{\kappa}\|^2 < (q-2)/N$ then $\Delta R(\boldsymbol{\mu}) > 0$ for all c such that

$$0 < c < 2c^* = \frac{2(q-2)}{N}\left(1-\left(\frac{N}{q-2}\right)^{1/2}\|\boldsymbol{\kappa}\|\right);$$

moreover, over this range of c, the right-hand side of (3.7) is maximized at $c = c^*$. Evaluating the right-hand side of (3.7) at $c = c^*$, we obtain

$$\max_{0\le c\le 2c^*} \Delta R(\boldsymbol{\mu}) \ge \frac{q-2}{N}\left(1-\left(\frac{N}{q-2}\right)^{1/2}\|\boldsymbol{\kappa}\|\right)^2.$$

This completes the proof of (1.5). □

Our first main result determines conditions under which the estimator (1.3) has lower risk than $\widehat{\boldsymbol{\mu}}$. As a consequence, we extend the classical result of James and Stein [24] to the setting of two-step, monotone incomplete samples. For the case in which $n = N$, the following result reduces immediately to the theorem of James-Stein.

Theorem 3.1. *Suppose that $p \ge 2$, $n \ge q+3$, $\boldsymbol{\Sigma} = \boldsymbol{I}_{p+q}$, and define*

$$c^* = \frac{p-2}{n} + \frac{q}{N}. \tag{3.8}$$

Then, with the loss function (1.2), $R(\boldsymbol{\mu};\widehat{\boldsymbol{\mu}}) > R\big(\boldsymbol{\mu};\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c)\big)$ for all $\boldsymbol{\mu}$ and all $c \in (0, 2c^*)$.

Proof. Since $p \ge 2$ and $q \ge 1$ then $c^* > 0$. We shall show that, for $0 < c < 2c^*$,

$$\Delta R(\boldsymbol{\mu}) = E\big[L(\boldsymbol{\mu},\widehat{\boldsymbol{\mu}}) - L\big(\boldsymbol{\mu},\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c)\big)\big],$$

the incremental risk in the estimator $\widehat{\boldsymbol{\mu}}$ over $\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c)$, is strictly positive for all $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Proceeding as at (3.1) and (3.2), we find that (3.3) remains valid here. Now let $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ denote the column vectors consisting of the first p and last q components, respectively, of $\boldsymbol{\nu}$, so that $\boldsymbol{\nu} = (\boldsymbol{\nu}_1', \boldsymbol{\nu}_2')'$; then,

$$(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu})'(\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}) \equiv (\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\mu}_{\boldsymbol{X}})'(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\nu}_1) + (\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\mu}_{\boldsymbol{Y}})'(\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2)$$

and

$$\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2 \equiv \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\nu}_1\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2\|^2. \tag{3.9}$$

Let

$$T_1 = \frac{1}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\mu}_{\boldsymbol{X}})'(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\nu}_1)$$

and

$$T_2 = \frac{1}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}(\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\mu}_{\boldsymbol{Y}})'(\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2),$$

so that

$$\Delta R(\boldsymbol{\mu}) = c\,E\Big\{2(T_1+T_2) - \frac{c}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}\Big\}$$

By (2.5), $\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}$ and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$ are independent; therefore,

$$E(T_2) = E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}} E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}} \frac{(\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\mu}_{\boldsymbol{Y}})'(\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2)}{\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2}. \tag{3.10}$$

Again by (2.5), $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} \sim N_q(\boldsymbol{\mu}_{\boldsymbol{Y}}, \frac{1}{N}\boldsymbol{I}_q)$; hence, in order to evaluate the inner expectation with respect to $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$ in (3.10), while holding $\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}$ fixed, we apply Stein's Lemma, equation (2.10), with $d \equiv q$, $\boldsymbol{Z} \equiv \widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$, $\boldsymbol{\theta} \equiv \boldsymbol{\mu}_{\boldsymbol{Y}}$, $\lambda^2 \equiv 1/N$, and

$$\boldsymbol{\Psi}(\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}) \equiv \frac{\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2}{\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2}.$$

Noting that $T_2 \equiv (\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\mu}_{\boldsymbol{Y}})'\boldsymbol{\Psi}(\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}})$, it follows from (2.10) that

$$\begin{aligned} E(T_2) &= N^{-1} E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}} E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}} \Bigg[\frac{q}{\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2} \\ &\qquad\qquad - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2}{(\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2)^2} \Bigg] \\ &= N^{-1} E \left[\frac{q}{\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^2} - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2}{\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^4} \right]. \end{aligned}$$

Next, we write

$$\begin{aligned} E(T_1) &= E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}} E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}} \frac{(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}})'(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1)}{\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^2} \\ &\equiv E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}} E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}} \frac{(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}})'(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1)}{\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2}. \end{aligned} \tag{3.11}$$

We now recall from Corollary 2.2 that $\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}|Q \sim N_p\big(\boldsymbol{\mu}_{\boldsymbol{X}}, n^{-1}(1+\bar{\tau}Q)\boldsymbol{I}_p\big)$. Rewriting (3.11) in the form

$$E(T_1) = E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}} E_Q E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}|Q} \frac{(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{X}})'(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1)}{\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2},$$

we evaluate the inner expectation with respect to $\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}|Q$ by applying Stein's Lemma, (2.10), with $d \equiv p$, $\boldsymbol{Z} \equiv \widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}$, $\theta \equiv \boldsymbol{\mu}_{\boldsymbol{X}}$, $\lambda^2 = n^{-1}(1+\bar{\tau}Q)$, and

$$\boldsymbol{\Psi}(\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}) \equiv \frac{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1}{\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2}.$$

Proceeding as before, we obtain

$$\begin{aligned} E(T_1) &= \frac{1}{n} E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}} E_Q E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}|Q} (1+\bar{\tau}Q) \Bigg[\frac{p}{\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2} \\ &\qquad\qquad - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2}{(\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2)^2} \Bigg] \\ &= \frac{1}{n} E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}} E_Q E_{\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}|Q} (1+\bar{\tau}Q) \left(\frac{p}{\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^2} - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2}{\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^4} \right). \end{aligned}$$

By (3.9),

$$\begin{aligned} \frac{p}{\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^2} - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2}{\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^4} &= \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^{-4} \left(p\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^2 - 2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2 \right) \\ &= \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}\|^{-4} \left((p-2)\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \boldsymbol{\nu}_1\|^2 + p|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \boldsymbol{\nu}_2\|^2 \right) \geq 0 \end{aligned}$$

for $p \geq 2$; hence

$$E(T_1) \geqslant \frac{1}{n} E\left(\frac{p}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2} - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\nu}_1\|^2}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^4}\right).$$

Therefore

$$\begin{aligned}
\Delta R(\boldsymbol{\mu}) &= c\,E\left(2(T_1+T_2) - \frac{c}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}\right) \\
&\geqslant c\,E\Bigg(2n^{-1}\left[\frac{p}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2} - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\nu}_1\|^2}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^4}\right] \\
&\qquad\qquad +2N^{-1}\left[\frac{q}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2} - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2\|^2}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^4}\right] - \frac{c}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}\Bigg) \\
&= c\,E\left(2\frac{n^{-1}p+N^{-1}q}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2} - 4\frac{n^{-1}\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\nu}_1\|^2 + N^{-1}\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2\|^2}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^4} - \frac{c}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}\right).
\end{aligned}$$

Since $n \leq N$ then

$$\begin{aligned}
n^{-1}\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\nu}_1\|^2 + N^{-1}\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2\|^2 &\leq n^{-1}(\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\boldsymbol{\nu}_1\|^2 + \|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\boldsymbol{\nu}_2\|^2) \\
&= n^{-1}\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2;
\end{aligned}$$

therefore

$$\begin{aligned}
\Delta R(\boldsymbol{\mu}) &\geqslant c\,E\left(\frac{2(n^{-1}p+N^{-1}q)}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2} - \frac{4n^{-1}\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^4} - \frac{c}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}\right) \\
&= c\big(2(n^{-1}p+N^{-1}q) - 4n^{-1} - c\big)\,E\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^{-2} \\
&\equiv c(2c^* - c)\,E\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^{-2}.
\end{aligned}$$

Therefore $\Delta R(\boldsymbol{\mu}) > 0$ for $0 < c < 2c^*$. Moreover, to maximize this lower bound, we choose $c = c^*$.

To complete the proof, we verify that $E\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^{-2} < \infty$ for $p+q > 2$. By Corollary 2.2, $(\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu})|Q \sim N_{p+q}(\boldsymbol{\mu}-\boldsymbol{\nu}, \boldsymbol{C}_Q)$. Since $n^{-1}(1+\bar{\tau}Q^2) \geq N^{-1}$ then the smallest eigenvalue of $\boldsymbol{C}_Q$ is N^{-1}. Hence, by Lemma 2.4,

$$E(\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^{-2}\big|Q) \leq \frac{N}{p+q-2} < \infty,$$

for $p+q > 2$, and the same conclusion holds for the unconditional expectation. $\square$

Analogous to the phenomenon that the classical, unrestricted James-Stein estimator is shown to be inadmissible by proving that it has larger risk than the positive-part James-Stein estimator, we shall show that the estimator $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ in (1.3) is inadmissible, having higher risk than its positive-part analog,

$$\boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c) = \left(1 - \frac{c}{\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}\|^2}\right)_+ (\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu}) + \boldsymbol{\nu}, \tag{3.12}$$

where, for $t \in \mathbb{R}$,

$$t_+ = \begin{cases} t, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{3.13}$$

denotes the *positive part* of t. Extending a result of Baranchik [5], the following result shows that the unrestricted shrinkage estimator (1.3) is dominated by its positive-part analog.

Theorem 3.2. *Let $p \geq 2$, $n \geq q+3$, and $\boldsymbol{\Sigma} = \boldsymbol{I}_{p+q}$. Under the loss function (1.2), the positive-part estimator (3.12) has lower risk than $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$:*

$$R\big(\boldsymbol{\mu}; \boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)\big) > R\big(\boldsymbol{\mu}; \boldsymbol{m}^{+}(\widehat{\boldsymbol{\mu}}, c)\big)$$

for all $\boldsymbol{\mu} \in \mathbb{R}^{p+q}$ and all $c \in (0, 2c^*)$, where c^* is defined in (3.8). Therefore $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ both are inadmissible.

Proof. Let $\boldsymbol{U} = \widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}$, $\boldsymbol{\kappa} = \boldsymbol{\mu} - \boldsymbol{\nu}$, and define $g(t) = 1 - ct^{-2}$, $t \in \mathbb{R}$, $t \neq 0$. Then

$$\begin{aligned}
&E\|\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c) - \boldsymbol{\mu}\|^2 - E\|\boldsymbol{m}^{+}(\widehat{\boldsymbol{\mu}}, c) - \boldsymbol{\mu}\|^2 \\
&\qquad = E\left(\|g(\|\boldsymbol{U}\|)\,\boldsymbol{U} - \boldsymbol{\kappa}\|^2 - \|g^{+}(\|\boldsymbol{U}\|)\,\boldsymbol{U} - \boldsymbol{\kappa}\|^2\right) \\
&\qquad = E\left([g(\|\boldsymbol{U}\|)]^2 - \left[g^{+}(\|\boldsymbol{U}\|)\right]^2\right)\|\boldsymbol{U}\|^2 + 2E\Big(g^{+}(\|\boldsymbol{U}\|) - g(\|\boldsymbol{U}\|)\Big)\boldsymbol{\kappa}'\boldsymbol{U}.
\end{aligned}$$

Clearly, $[g(t)]^2 - [g^{+}(t)]^2 \geqslant 0$ for all t, and therefore

$$E\left([g(\|\boldsymbol{U}\|)]^2 - \left[g^{+}(\|\boldsymbol{U}\|)\right]^2\right)\|\boldsymbol{U}\|^2 \geqslant 0,$$

so we have

$$E\|\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c) - \boldsymbol{\mu}\|^2 - E\|\boldsymbol{m}^{+}(\widehat{\boldsymbol{\mu}}, c) - \boldsymbol{\mu}\|^2 \geq 2E\Big(g^{+}(\|\boldsymbol{U}\|) - g(\|\boldsymbol{U}\|)\Big)\boldsymbol{\kappa}'\boldsymbol{U}. \tag{3.14}$$

Denote by $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$ the vectors containing the first p and last q components, respectively, of $\boldsymbol{\kappa}$, so that $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1', \boldsymbol{\kappa}_2')'$. Let $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ be, respectively, $p \times p$ and $q \times q$, orthogonal matrices such that

$$\boldsymbol{H}_1'\boldsymbol{\kappa}_1 = \|\boldsymbol{\kappa}_1\|(1, \underbrace{0, \ldots, 0}_{p-1})'$$

and

$$\boldsymbol{H}_2'\boldsymbol{\kappa}_2 = \|\boldsymbol{\kappa}_2\|(\underbrace{0, \ldots, 0}_{q-1}, 1)'.$$

Denote by $\{\mathbf{e}_1, \ldots, \mathbf{e}_{p+q}\}$ the standard basis for $\mathbb{R}^{p+q}$; then it follows from above that

$$\begin{pmatrix} \boldsymbol{H}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2' \end{pmatrix}\boldsymbol{\kappa} = \|\boldsymbol{\kappa}_1\|\mathbf{e}_1 + \|\boldsymbol{\kappa}_2\|\mathbf{e}_{p+q}.$$

Since $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ both are orthogonal then

$$\begin{pmatrix} \boldsymbol{H}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2' \end{pmatrix}\boldsymbol{V}_1 \overset{\mathcal{L}}{=} \boldsymbol{V}_1, \qquad \begin{pmatrix} \boldsymbol{H}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2' \end{pmatrix}\begin{pmatrix} \boldsymbol{V}_2 \\ \mathbf{0} \end{pmatrix} \overset{\mathcal{L}}{=} \begin{pmatrix} \boldsymbol{V}_2 \\ \mathbf{0} \end{pmatrix},$$

and this orthogonal transformation also preserves the independence of $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$. Therefore

$$\begin{aligned}
\begin{pmatrix} \boldsymbol{H}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2' \end{pmatrix}\boldsymbol{U} &\overset{\mathcal{L}}{=} \begin{pmatrix} \boldsymbol{H}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2' \end{pmatrix}\left[\boldsymbol{\kappa} + \boldsymbol{V}_1 + \sqrt{\frac{\bar{\tau}Q}{n}}\begin{pmatrix} \boldsymbol{V}_2 \\ \mathbf{0} \end{pmatrix}\right] \\
&= \begin{pmatrix} \boldsymbol{H}_1'\boldsymbol{\kappa}_1 \\ \boldsymbol{H}_2'\boldsymbol{\kappa}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{H}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2' \end{pmatrix}\boldsymbol{V}_1 + \sqrt{\frac{\bar{\tau}Q}{n}}\begin{pmatrix} \boldsymbol{H}_1'\boldsymbol{V}_2 \\ \mathbf{0} \end{pmatrix} \\
&\overset{\mathcal{L}}{=} \|\boldsymbol{\kappa}_1\|\mathbf{e}_1 + \|\boldsymbol{\kappa}_2\|\mathbf{e}_{p+q} + \boldsymbol{V}_1 + \sqrt{\frac{\bar{\tau}Q}{n}}\begin{pmatrix} \boldsymbol{V}_2 \\ \mathbf{0} \end{pmatrix}.
\end{aligned} \tag{3.15}$$

Letting $\widetilde{\boldsymbol{U}} = \begin{pmatrix} \boldsymbol{H}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2' \end{pmatrix} \boldsymbol{U}$, and noting that $\|\widetilde{\boldsymbol{U}}\| = \|\boldsymbol{U}|$, it follows that the right-hand side of (3.14) equals

$$
\begin{aligned}
E\Big[g^{+}(\|\widetilde{\boldsymbol{U}}\|) - g(\|\widetilde{\boldsymbol{U}}\|)\Big]\, \boldsymbol{\kappa}' \begin{pmatrix} \boldsymbol{H}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2 \end{pmatrix} \widetilde{\boldsymbol{U}} \\
&= E\Big[g^{+}(\|\widetilde{\boldsymbol{U}}\|) - g(\|\widetilde{\boldsymbol{U}}\|)\Big] \left(\begin{pmatrix} \boldsymbol{H}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2' \end{pmatrix} \boldsymbol{\kappa} \right)' \widetilde{\boldsymbol{U}} \\
&= E\Big[g^{+}(\|\widetilde{\boldsymbol{U}}\|) - g(\|\widetilde{\boldsymbol{U}}\|)\Big] (\|\boldsymbol{\kappa}_1\|\mathbf{e}_1 + \|\boldsymbol{\kappa}_2\|\mathbf{e}_{p+q})' \widetilde{\boldsymbol{U}} \\
&= E\Big[g^{+}(\|\widetilde{\boldsymbol{U}}\|) - g(\|\widetilde{\boldsymbol{U}}\|)\Big] (\|\boldsymbol{\kappa}_1\|\widetilde{U}_1 + \|\boldsymbol{\kappa}_2\|\widetilde{U}_{p+q}), \qquad (3.16)
\end{aligned}
$$

where $\widetilde{U}_1, \ldots, \widetilde{U}_{p+q}$ are the components of $\widetilde{\boldsymbol{U}}$.

By (3.15) and Corollary 2.2, $\widetilde{\boldsymbol{U}}|Q \sim N_{p+q}(\|\boldsymbol{\kappa}_1\|\mathbf{e}_1 + \|\boldsymbol{\kappa}_2\|\mathbf{e}_{p+q}, \boldsymbol{C}_Q)$. Since $\boldsymbol{C}_Q$ is diagonal then, conditional on Q, the variables $\widetilde{U}_1, \ldots, \widetilde{U}_{p+q}$ are mutually independent; therefore

$$
E\Big[g^{+}(\|\widetilde{\boldsymbol{U}}\|) - g(\|\widetilde{\boldsymbol{U}}\|)\Big] \widetilde{U}_1 = E_Q\, E_{\widetilde{U}_2,\ldots,\widetilde{U}_{p+q}|Q}\, E_{\widetilde{U}_1|Q} \Big[g^{+}(\|\widetilde{\boldsymbol{U}}\|) - g(\|\widetilde{\boldsymbol{U}}\|)\Big] \widetilde{U}_1. \qquad (3.17)
$$

For fixed $\widetilde{U}_2, \ldots, \widetilde{U}_{p+q}$, define $h : \mathbb{R}_+ \to \mathbb{R}$ by

$$
h(t) = g\Big(\sqrt{t^2 + \widetilde{U}_2^2 + \cdots + \widetilde{U}_{p+q}^2}\Big);
$$

then, $h(\widetilde{U}_1) \equiv g(\|\widetilde{\boldsymbol{U}}\|)$ and $h^{+}(\widetilde{U}_1) \equiv g^{+}(\|\widetilde{\boldsymbol{U}}\|)$. By Lemma 3.5.2 of Anderson [2], p. 96, we recall that if Z is a one-dimensional normal random variable then, for any function $h_1 : \mathbb{R} \to \mathbb{R}$, $E[h_1^{+}(Z) - h_1(Z)]Z$, whenever it exists, has the same sign as $E(Z)$. Applying this result to $Z \equiv \widetilde{U}_1|Q \sim N_1(\|\boldsymbol{\kappa}_1\|, \frac{1}{n}(1+\bar{\tau}Q))$ and $h_1 \equiv h$, and bearing in mind that $E(\widetilde{U}_1|Q) = \|\boldsymbol{\kappa}_1\| \geq 0$, we obtain

$$
0 \leq E_{\widetilde{U}_1|Q}\Big[h^{+}(\widetilde{U}_1) - h(\widetilde{U}_1)\Big] \widetilde{U}_1 \equiv E_{\widetilde{U}_1|Q}\Big[g^{+}(\|\widetilde{\boldsymbol{U}}\|) - g(\|\widetilde{\boldsymbol{U}}\|)\Big] \widetilde{U}_1,
$$

for fixed $\widetilde{U}_2, \ldots, \widetilde{U}_{p+q}$. Inserting this inequality at (3.17), we obtain

$$
E\Big[g^{+}(\|\widetilde{\boldsymbol{U}}\|) - g(\|\widetilde{\boldsymbol{U}}\|)\Big] \widetilde{U}_1 \geq 0. \qquad (3.18)
$$

By a similar argument, we deduce also that

$$
E\Big[g^{+}(\|\widetilde{\boldsymbol{U}}\|) - g(\|\widetilde{\boldsymbol{U}}\|)\Big] \widetilde{U}_{p+q} \geq 0. \qquad (3.19)
$$

It now follows from (3.18) and (3.19) that (3.16), and hence (3.14), is nonnegative. $\square$

4 The case of block-diagonal covariance matrices

For the case in which $\boldsymbol{\Sigma}$ is arbitrary and known, many results are available in the literature in the classical case in which the sample is complete; see *e.g.*, Anderson [2], Berger [8], and Brown [14] for citations of the original sources for those results. In the setting of two-step monotone incomplete normal samples, the derivation of improved estimation through shrinkage estimators has proved to be considerably difficult for the case in which $\boldsymbol{\Sigma}$ is arbitrary. If $\boldsymbol{\Sigma}$ is diagonal or block-diagonal, however, then we have derived results that show that shrinkage lowers the risk of $\widehat{\boldsymbol{\mu}}$. Let us first consider the diagonal case.

Suppose that $\boldsymbol{\Sigma}$ is a known diagonal matrix and the loss function L is the same as in (1.2), *viz.*, $L(\boldsymbol{\mu}, \boldsymbol{m}) = \|\boldsymbol{m} - \boldsymbol{\mu}\|^2$. In stating the following result, we use the notation

$$\eta_* = \left(1 + \frac{q(N-n)}{(n-q-2)N}\right)^{-1}, \tag{4.1}$$

We also denote by $\boldsymbol{\Sigma}_*$ the covariance matrix of $\widehat{\boldsymbol{\mu}}$, an explicit formula for which is provided in (2.6). As background for the following result, we refer to Berger [8], pp. 363–369 for citations to the classical literature on results of this type.

Theorem 4.1. *Let $p \ge 2$, $n \ge q+3$, and let $r(t)$, $t \ge 0$ be a nondecreasing, differentiable function such that $0 \le r(t) \le 2(\eta_* p + q - 2)$. Let $\boldsymbol{\Sigma}$ be a known, diagonal matrix and $\boldsymbol{\nu} \in \mathbb{R}^{p+q}$. Then, with the loss function (1.2), the estimator*

$$\boldsymbol{m}(\widehat{\boldsymbol{\mu}}) = \left(\boldsymbol{I}_{p+q} - \frac{r\big((\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})'{\boldsymbol{\Sigma}_*}^{-2}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})\big)}{(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})'{\boldsymbol{\Sigma}_*}^{-2}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})}{\boldsymbol{\Sigma}_*}^{-1}\right)(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}) + \boldsymbol{\nu} \tag{4.2}$$

has smaller risk than $\widehat{\boldsymbol{\mu}}$.

Proof. We show, first, that $\eta_* p + q - 2 > 0$. This inequality obviously holds if $q \ge 2$; and if $q = 1$ then it follows from

$$\begin{aligned} p + \eta_*^{-1}(q-2) &\equiv p - \left(1 + \frac{\bar{\tau}}{n-3}\right) \\ &\ge 2 - \left(1 + \frac{\bar{\tau}}{n-3}\right) = \frac{(n-4)N + n}{(n-3)N} > 0. \end{aligned}$$

Now let $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_{11}, \cdots, \sigma_{p+q,p+q})$. Then

$$\boldsymbol{\Omega} = (\omega_{ij}) = \frac{1}{N}\boldsymbol{\Sigma} + \frac{\bar{\tau}}{n}\begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix},$$

so that $\omega_{ij} = 0$, $i \ne j$ and

$$\omega_{ii} = \begin{cases} \frac{1}{n}\sigma_{ii}, & i = 1, \ldots, p \\ \frac{1}{N}\sigma_{ii}, & i = p+1, \ldots, p+q. \end{cases}$$

By (2.6), we also have $\boldsymbol{\Sigma}_* = (\sigma^*_{ij})$ where $\sigma^*_{ij} = 0$, $i \ne j$ and

$$\sigma^*_{ii} = \begin{cases} \left(\frac{1}{n} + \frac{q\bar{\tau}}{n(n-q-2)}\right)\sigma_{ii}, & i = 1, \ldots, p \\ \frac{1}{N}\sigma_{ii}, & i = p+1, \ldots, p+q. \end{cases} \tag{4.3}$$

Let $\boldsymbol{U} = \widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}$ and $\boldsymbol{\kappa} = \boldsymbol{\mu} - \boldsymbol{\nu}$, so that

$$\boldsymbol{U}|Q = (\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})|Q \sim N_{p+q}\left(\boldsymbol{\kappa}, \boldsymbol{\Omega} + \frac{\bar{\tau}}{n}Q\begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\right).$$

It is elementary that $L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) = L(\boldsymbol{\kappa}, \boldsymbol{U})$ and $L(\boldsymbol{\mu}, \boldsymbol{m}(\widehat{\boldsymbol{\mu}})) = L(\boldsymbol{\kappa}, \boldsymbol{m}_*(\boldsymbol{U}))$, where

$$\boldsymbol{m}_*(\boldsymbol{U}) = \Big(\boldsymbol{I}_{p+q} - \frac{r(\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U}}\boldsymbol{\Sigma}_*{}^{-1}\Big)\boldsymbol{U}.$$

Denote by $U_1, \ldots, U_{p+q}$ the components of $\boldsymbol{U}$; then $\Delta R(\boldsymbol{\mu}) = E[L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) - L(\boldsymbol{\mu}, \boldsymbol{m}(\widehat{\boldsymbol{\mu}}))]$, the difference in risks between $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{m}(\widehat{\boldsymbol{\mu}})$, equals

$$\begin{aligned} &E\{\|\boldsymbol{U} - \boldsymbol{\kappa}\|^2 - \|\boldsymbol{m}_*(\boldsymbol{U}) - \boldsymbol{\kappa}\|^2\} \\ &\qquad = E\Big\{2\frac{r(\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U}} \sum_{i=1}^{p+q} \frac{U_i(U_i - \mu_i^*)}{\sigma_{ii}^*} - \frac{r^2(\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U}}\Big\}. \end{aligned} \tag{4.4}$$

We now apply Stein's Lemma, (2.9), with $d \equiv p + q$, $\boldsymbol{Z} \equiv \boldsymbol{U}|Q$, $\boldsymbol{\theta} \equiv \boldsymbol{\kappa}$, $\Lambda \equiv \mathrm{Cov}(\boldsymbol{U}|Q)$, and

$$\psi(\boldsymbol{U}) = \frac{r(\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U}} U_i.$$

Accordingly, we obtain

$$E(U_i - \mu_i^*)\psi(\boldsymbol{U})|Q = \begin{cases} \frac{1}{n}\sigma_{ii}(1 + \bar{\tau}Q)\, E_{\boldsymbol{U}|Q}\, \psi_i(\boldsymbol{U}), & 1 \le i \le p \\ \frac{1}{N}\sigma_{ii}\, E_{\boldsymbol{U}|Q}\, \psi_i(\boldsymbol{U}), & p + 1 \le i \le p + q, \end{cases} \tag{4.5}$$

where

$$\psi_i(\boldsymbol{U}) = \frac{\partial \psi(\boldsymbol{U})}{\partial U_i} = \frac{r(\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U}} + \frac{2r'(\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U}} \frac{U_i^2}{\sigma_{ii}^{*\,2}} - \frac{2r(\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U})}{(\boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U})^2} \frac{U_i^2}{\sigma_{ii}^{*\,2}}.$$

Let $W \equiv \boldsymbol{U}'\boldsymbol{\Sigma}_*{}^{-2}\boldsymbol{U}$; substituting (4.5) into the right-hand side of (4.4), we obtain

$$\begin{aligned} \Delta R(\boldsymbol{\mu}) = E\Big\{&2(1 + \bar{\tau}Q) \sum_{i=1}^{p} \frac{\sigma_{ii}}{n\sigma_{ii}^*}\Big(\frac{r(W)}{W} + \frac{2r'(W)}{W}\frac{U_i^2}{\sigma_{ii}^{*\,2}} - \frac{2r(W)}{W^2}\frac{U_i^2}{\sigma_{ii}^{*\,2}}\Big) \\ &+ 2\sum_{i=p+1}^{p+q} \frac{\sigma_{ii}}{N\sigma_{ii}^*}\Big(\frac{r(W)}{W} + \frac{2r'(W)}{W}\frac{U_i^2}{\sigma_{ii}^{*\,2}} - \frac{2r(W)}{W^2}\frac{U_i^2}{\sigma_{ii}^{*\,2}}\Big) - \frac{r^2(W)}{W}\Big\}. \end{aligned}$$

We introduce some temporary notation in order to simplify this expression. Let

$$W_1 = \sum_{i=1}^{p} \frac{U_i^2}{\sigma_{ii}^{*\,2}}, \quad W_2 = \sum_{i=p+1}^{p+q} \frac{U_i^2}{\sigma_{ii}^{*\,2}},$$

so that $W = W_1 + W_2$. Also, let $\eta = \eta_*(1 + \bar{\tau}Q)$ where η_* is defined in (4.1); then, clearly, $\eta \ge \eta_*$. By (4.3),

$$\frac{\sigma_{ii}}{\sigma_{ii}^*} = \begin{cases} n\,\eta_*, & i = 1, \ldots, p, \\ N, & i = p + 1, \ldots, p + q. \end{cases}$$

Therefore

$$
\begin{aligned}
\Delta R(\boldsymbol{\mu}) &= E\Big\{2\eta \sum_{i=1}^{p} \Big(\frac{r(W)}{W} + \frac{2r'(W)}{W}\frac{U_i^2}{\sigma_{ii}^{*\,2}} - \frac{2r(W)}{W^2}\frac{U_i^2}{\sigma_{ii}^{*\,2}}\Big) \\
&\qquad + 2\sum_{i=p+1}^{p+q} \Big(\frac{r(W)}{W} + \frac{2r'(W)}{W}\frac{U_i^2}{\sigma_{ii}^{*\,2}} - \frac{2r(W)}{W^2}\frac{U_i^2}{\sigma_{ii}^{*\,2}}\Big) - \frac{r^2(W)}{W}\Big\} \\
&= E\Big\{2\eta\Big(p\frac{r(W)}{W} + \frac{2r'(W)}{W}W_1 - \frac{2r(W)}{W^2}W_1\Big) \\
&\qquad + 2\Big(q\frac{r(W)}{W} + \frac{2r'(W)}{W}W_2 - \frac{2r(W)}{W^2}W_2\Big) - \frac{r^2(W)}{W}\Big\} \\
&\equiv E\big\{W^{-1}F(\eta, W_1, W_2)\big\},
\end{aligned} \tag{4.6}
$$

where

$$
F(\eta, w_1, w_2) = 2\Big(\eta p + q - \frac{2(\eta w_1 + w_2)}{w}\Big)r(w) + 4(\eta w_1 + w_2)r'(w) - r^2(w),
$$

$w_1, w_2 \ge 0$, $w = w_1 + w_2$.

Suppose that $\eta \ge 1$; since $0 \le w_2 \le w$ then

$$
\eta w_1 + w_2 = \eta(w - w_2) + w_2 = \eta w + (1-\eta)w_2 \le \eta w;
$$

consequently,

$$
\begin{aligned}
F(\eta, w_1, w_2) &\ge 2\big(\eta(p-2) + q\big)r(w) + 4(\eta w_1 + w_2)r'(w) - r^2(w) \\
&\ge 2\big(\eta^*(p-2) + q\big)r(w) + 4(\eta w_1 + w_2)r'(w) - r^2(w),
\end{aligned}
$$

the last inequality being due to the fact that $\eta \ge \eta^*$. Since $r(\cdot)$ is nondecreasing then $r'(w) \ge 0$; therefore

$$
\begin{aligned}
F(\eta, w_1, w_2) &\ge 2\big(\eta^*(p-2) + q\big)r(w) - r^2(w) \\
&\equiv \big(4(1-\eta^*) + 2(\eta^* p + q - 2) - r(w)\big)r(w).
\end{aligned}
$$

Since $\eta^* < 1$ for $n \ge q + 3$ and, by assumption, $r(w) \le 2(\eta^* p + q - 2)$, it follows that $F(\eta, w_1, w_2) > 0$ for all w if $\eta > 1$.

Suppose next that $\eta < 1$; then $\eta w_1 + w_2 < w_1 + w_2 = w$, and therefore

$$
\begin{aligned}
F(\eta, w_1, w_2) &\ge 2(\eta p + q - 2)r(w) + 4(\eta w_1 + w_2)r'(w) - r^2(w) \\
&\ge \big(2(\eta^* p + q - 2) - r(w)\big)r(w)
\end{aligned}
$$

since $\eta \ge \eta^*$ and $r'(w) \ge 0$. As before, we apply the assumption $r(w) \le 2(\eta^* p + q - 2)$ to deduce that $F(\eta, w_1, w_2) \ge 0$ for all w if $\eta \le 1$.

Returning to (4.6), we apply the Law of Total Probability to write

$$
\begin{aligned}
E\big\{W^{-1}F(\eta, W_1, W_2)\big\} \;&=\; E\big\{W^{-1}F(\eta, W_1, W_2)|\eta > 1\big\} \cdot P(\eta > 1) \\
&\qquad + E\big\{W^{-1}F(\eta, W_1, W_2)|\eta \le 1\big\} \cdot P(\eta \le 1)\big\},
\end{aligned}
$$

proving that $\Delta R(\boldsymbol{\mu}) \ge 0$. $\square$

We have also obtained results for the case in which $\boldsymbol{\Sigma}$ is block-diagonal,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \tag{4.7}$$

and the loss function is

$$L(\boldsymbol{\mu}, \boldsymbol{m}) = (\boldsymbol{m} - \boldsymbol{\mu})' \boldsymbol{M} (\boldsymbol{m} - \boldsymbol{\mu}), \tag{4.8}$$

where $\boldsymbol{M}$ is block-diagonal,

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{M}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{M}_{22} \end{pmatrix}, \tag{4.9}$$

and $\boldsymbol{M}_{11}$ and $\boldsymbol{M}_{22}$ are arbitrary $p \times p$ and $q \times q$ positive definite matrices, respectively. Then, we have the following result.

Theorem 4.2. *Suppose that $\boldsymbol{\Sigma}$ in (4.7) is known and that the loss function is (4.8) where $\boldsymbol{M}$ is given by (4.9). Let $\boldsymbol{\nu} \in \mathbb{R}^{p+q}$, and let $r(t)$, $t \ge 0$ be a non-decreasing, differentiable function such that $0 \le r(t) \le 2(\eta_* p + q - 2)$ for all t, where η_* is given in (4.1). Then, for $p \ge 2$ and $n \ge q + 3$, the estimator*

$$\boldsymbol{m}(\widehat{\boldsymbol{\mu}}) = \Big(\boldsymbol{I}_{p+q} - \frac{r\big((\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})' {\boldsymbol{\Sigma}_*}^{-1} \boldsymbol{M}^{-1} {\boldsymbol{\Sigma}_*}^{-1} (\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})\big)}{(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})' {\boldsymbol{\Sigma}_*}^{-1} \boldsymbol{M}^{-1} {\boldsymbol{\Sigma}_*}^{-1} (\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})} \boldsymbol{M}^{-1} {\boldsymbol{\Sigma}_*}^{-1} \Big) (\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}) + \boldsymbol{\nu} \tag{4.10}$$

has smaller risk than $\widehat{\boldsymbol{\mu}}$.

Proof. There exists a nonsingular, $p \times p$ matrix $\boldsymbol{C}_{11}$ such that $\boldsymbol{C}'_{11} \boldsymbol{M}_{11} \boldsymbol{C}_{11} = \boldsymbol{I}_p$ and $\boldsymbol{\Sigma}_{11} = \boldsymbol{C}_{11} \boldsymbol{\Delta}_1 \boldsymbol{C}'_{11}$, where $\boldsymbol{\Delta}_1$ is a diagonal matrix with positive diagonal entries. Similarly, there exists a nonsingular, $q \times q$ matrix $\boldsymbol{C}_{22}$ satisfying $\boldsymbol{C}'_{22} \boldsymbol{M}_{22} \boldsymbol{C}_{22} = \boldsymbol{I}_q$ and $\boldsymbol{\Sigma}_{22} = \boldsymbol{C}_{22} \boldsymbol{\Delta}_2 \boldsymbol{C}'_{22}$ where $\boldsymbol{\Delta}_2$ is a diagonal matrix with positive diagonal entries. Define the block-diagonal matrix

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{C}_{22} \end{pmatrix},$$

introduce the transformation

$$\boldsymbol{U} = \boldsymbol{C}^{-1} (\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}), \tag{4.11}$$

and let $\boldsymbol{\mu}_* = \boldsymbol{C}^{-1}(\boldsymbol{\mu} - \boldsymbol{\nu})$. Recalling that $\widehat{\boldsymbol{\mu}}$ is unbiased, and applying the formula (2.6) for the covariance matrix of $\widehat{\boldsymbol{\mu}}$, we obtain $E(\boldsymbol{U}) = \boldsymbol{\mu}_*$ and $\mathrm{Cov}(\boldsymbol{U}) = \boldsymbol{\Sigma}_{**}$ where

$$\boldsymbol{\Sigma}_{**} = \boldsymbol{C}^{-1} \boldsymbol{\Sigma}_* \boldsymbol{C}^{-1\prime} = \begin{pmatrix} \frac{1}{n} \boldsymbol{\Delta}_1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{N} \boldsymbol{\Delta}_2 \end{pmatrix} + \frac{q \bar{\tau}}{n(n-q-2)} \begin{pmatrix} \boldsymbol{\Delta}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

a diagonal matrix.

Let $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$, respectively, denote the vectors of the first p and last q components of $\boldsymbol{U}$. Then it follows from (2.5) and (4.11) that

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix} \stackrel{\mathcal{L}}{=} \boldsymbol{\mu}_* + \widetilde{\boldsymbol{V}}_1 + \sqrt{\frac{\bar{\tau} Q}{n}} \begin{pmatrix} \widetilde{\boldsymbol{V}}_2 \\ \mathbf{0} \end{pmatrix},$$

where $\widetilde{\boldsymbol{V}}_1 \sim N_{p+q}(\mathbf{0}, \widetilde{\boldsymbol{\Omega}})$ and $\widetilde{\boldsymbol{V}}_2 \sim N_p(\mathbf{0}, \boldsymbol{\Delta}_1)$ with

$$\widetilde{\boldsymbol{\Omega}} = \boldsymbol{C}^{-1} \boldsymbol{\Omega} \boldsymbol{C}^{-1\prime} = \begin{pmatrix} \frac{1}{n} \boldsymbol{\Delta}_1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{N} \boldsymbol{\Delta}_2 \end{pmatrix}.$$

Under the transformation (4.11), the statistic $(\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu})'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{M}^{-1}\boldsymbol{\Sigma}_*^{-1}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\nu})$ is transformed to

$$\begin{aligned}\boldsymbol{U}'\boldsymbol{C}'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{M}^{-1}\boldsymbol{\Sigma}_*^{-1}\boldsymbol{C}\boldsymbol{U} &= \boldsymbol{U}'(\boldsymbol{C}'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{C})(\boldsymbol{C}^{-1}\boldsymbol{M}^{-1}\boldsymbol{C}'^{-1})(\boldsymbol{C}'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{C})\boldsymbol{U}\\ &= \boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-1}(\boldsymbol{C}'\boldsymbol{M}\boldsymbol{C})^{-1}\boldsymbol{\Sigma}_{**}^{-1}\boldsymbol{U}\\ &= \boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-2}\boldsymbol{U},\end{aligned}$$

because $\boldsymbol{C}'\boldsymbol{M}\boldsymbol{C} = \boldsymbol{I}_{p+q}$. Also, the estimator $\boldsymbol{m}(\widehat{\boldsymbol{\mu}})$ in (4.10) is transformed to $\boldsymbol{m}_*(\boldsymbol{U})$, where

$$\boldsymbol{m}_*(\boldsymbol{U}) = \Big(\boldsymbol{I}_{p+q} - \frac{r(\boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-2}\boldsymbol{U}}\boldsymbol{M}^{-1}\boldsymbol{\Sigma}_*^{-1}\Big)\boldsymbol{C}\boldsymbol{U} + \boldsymbol{\nu}.$$

Since $\boldsymbol{M}^{-1}\boldsymbol{\Sigma}_*^{-1}\boldsymbol{C} \equiv \boldsymbol{C}(\boldsymbol{C}'\boldsymbol{M}\boldsymbol{C})^{-1}(\boldsymbol{C}'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{C}) = \boldsymbol{C}\boldsymbol{\Sigma}_{**}^{-1}$ then $\boldsymbol{M}^{-1}\boldsymbol{\Sigma}_*^{-1} = \boldsymbol{C}\boldsymbol{\Sigma}_{**}^{-1}\boldsymbol{C}^{-1}$; also, $\boldsymbol{\nu} = \boldsymbol{\mu} - \boldsymbol{C}\boldsymbol{\mu}_*$. Therefore,

$$\begin{aligned}\boldsymbol{m}_*(\boldsymbol{U}) - \boldsymbol{\mu} &= \Big(\boldsymbol{I}_{p+q} - \frac{r(\boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-2}\boldsymbol{U}}\boldsymbol{C}\boldsymbol{\Sigma}_{**}^{-1}\boldsymbol{C}^{-1}\Big)\boldsymbol{C}\boldsymbol{U} - \boldsymbol{C}\boldsymbol{\mu}_*\\ &= \boldsymbol{C}\bigg(\Big(\boldsymbol{I}_{p+q} - \frac{r(\boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-2}\boldsymbol{U}}\boldsymbol{\Sigma}_{**}^{-1}\Big)\boldsymbol{U} - \boldsymbol{\mu}_*\bigg)\\ &\equiv \boldsymbol{C}\big(\boldsymbol{m}_{**}(\boldsymbol{U}) - \boldsymbol{\mu}_*\big),\end{aligned}$$

where

$$\boldsymbol{m}_{**}(\boldsymbol{U}) = \Big(\boldsymbol{I}_{p+q} - \frac{r(\boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-2}\boldsymbol{U})}{\boldsymbol{U}'\boldsymbol{\Sigma}_{**}^{-2}\boldsymbol{U}}\boldsymbol{\Sigma}_{**}^{-1}\Big)\boldsymbol{U}.$$

Therefore,

$$\begin{aligned}L(\boldsymbol{\mu}, \boldsymbol{m}(\widehat{\boldsymbol{\mu}})) &= (\boldsymbol{m}(\widehat{\boldsymbol{\mu}}) - \boldsymbol{\mu})'\boldsymbol{M}(\boldsymbol{m}(\widehat{\boldsymbol{\mu}}) - \boldsymbol{\mu})\\ &= (\boldsymbol{C}(\boldsymbol{m}_{**}(\boldsymbol{U}) - \boldsymbol{\mu}_*))'\boldsymbol{M}(\boldsymbol{C}(\boldsymbol{m}_{**}(\boldsymbol{U}) - \boldsymbol{\mu}_*))\\ &= (\boldsymbol{m}_{**}(\boldsymbol{U}) - \boldsymbol{\mu}_*)'\boldsymbol{C}'\boldsymbol{M}\boldsymbol{C}(\boldsymbol{m}_{**}(\boldsymbol{U}) - \boldsymbol{\mu}_*)\\ &= \|\boldsymbol{\mu}_* - \boldsymbol{m}_{**}(\boldsymbol{U})\|^2\\ &\equiv L(\boldsymbol{\mu}_*, \boldsymbol{m}_{**}(\boldsymbol{U})).\end{aligned}$$

This reduces the problem to the case in which $\boldsymbol{\Sigma}$ is diagonal and the loss function is the squared error loss in (1.2), so the conclusion follows from Theorem 4.1. □

With regard to general non-radial loss functions, we consider the case in which the loss function is again of the form (4.8), where

$$\boldsymbol{M} = \begin{pmatrix}\boldsymbol{M}_{11} & \boldsymbol{M}_{12}\\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22}\end{pmatrix} \tag{4.12}$$

is a general $(p+q)\times(p+q)$ positive definite symmetric, with $\boldsymbol{M}_{11}$ and $\boldsymbol{M}_{22}$ being $p\times p$ and $q\times q$ matrices, respectively. Then we have the following result.

Theorem 4.3. *Let $r(t)$, $t \ge 0$, be a nondecreasing, differentiable function such that $0 \le r(t) \le 2(\eta_* p + q - 2)$ where η_* is given in (4.1). Let (4.8) be the loss function where $\boldsymbol{M}$ is of* *the form (4.12), and suppose that* $\boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1} = -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. *Then, for* $\boldsymbol{\nu} \in \mathbb{R}^{p+q}$, $p \geq 2$, *and* $n > q + 3$, *the estimator*

$$\boldsymbol{m}(\widehat{\boldsymbol{\mu}}) = \left(\boldsymbol{I}_{p+q} - \frac{r\big((\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})'{\boldsymbol{\Sigma}_*}^{-1}\boldsymbol{M}^{-1}{\boldsymbol{\Sigma}_*}^{-1}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})\big)}{(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})'{\boldsymbol{\Sigma}_*}^{-1}\boldsymbol{M}^{-1}{\boldsymbol{\Sigma}_*}^{-1}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})}\boldsymbol{M}^{-1}{\boldsymbol{\Sigma}_*}^{-1}\right)(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}) + \boldsymbol{\nu}$$

has smaller risk than $\widehat{\boldsymbol{\mu}}$. *In particular, this result holds for the case in which* $\boldsymbol{M} = \boldsymbol{\Sigma}^{-1}$.

Proof. By (2.5), we have

$$\boldsymbol{\Sigma}_* = \mathrm{Cov}(\widehat{\boldsymbol{\mu}}) = \boldsymbol{\Omega} + \frac{q\bar{\tau}}{n(n-q-2)}\begin{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}.$$

Consider the estimators $\boldsymbol{m}_1 \equiv \widehat{\boldsymbol{\mu}}$ and $\boldsymbol{m}_2 \equiv \boldsymbol{m}(\widehat{\boldsymbol{\mu}})$, where

$$\boldsymbol{m}(\widehat{\boldsymbol{\mu}}) = \Big(\boldsymbol{I}_{p+q} - \frac{r\left((\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{M}^{-1}\boldsymbol{\Sigma}_*^{-1}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})\right)}{(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})'\boldsymbol{\Sigma}_*^{-1}\boldsymbol{M}^{-1}\boldsymbol{\Sigma}_*^{-1}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})}\boldsymbol{M}^{-1}\boldsymbol{\Sigma}_*^{-1}\Big)(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}) + \boldsymbol{\nu}.$$

We make the transformation $\boldsymbol{U} = \boldsymbol{C}\widehat{\boldsymbol{\mu}}$ where

$$\boldsymbol{C} = \begin{pmatrix}\boldsymbol{I}_p & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ \boldsymbol{0} & \boldsymbol{I}_q\end{pmatrix};$$

it is well-known that

$$\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}' = \begin{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22}\end{pmatrix}.$$

By (2.5),

$$\boldsymbol{U} \stackrel{\mathcal{L}}{=} \widetilde{\boldsymbol{\mu}} + \widetilde{\boldsymbol{V}}_1 + \sqrt{\frac{\bar{\tau}Q}{n}}\begin{pmatrix}\boldsymbol{V}_2\\ \boldsymbol{0}\end{pmatrix}$$

where $\widetilde{\boldsymbol{\mu}} = \boldsymbol{C}\boldsymbol{\mu}$ and $\widetilde{\boldsymbol{V}}_1 = \boldsymbol{C}\boldsymbol{V} \sim N_{p+q}(\boldsymbol{0}, \widetilde{\boldsymbol{\Omega}})$ with

$$\widetilde{\boldsymbol{\Omega}} = \boldsymbol{C}\boldsymbol{\Omega}\boldsymbol{C}' = \begin{pmatrix}\frac{1}{n}\boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0}\\ \boldsymbol{0} & \frac{1}{N}\boldsymbol{\Sigma}_{22}\end{pmatrix}.$$

Let $\widetilde{\boldsymbol{\nu}} = \boldsymbol{C}\boldsymbol{\nu}$, $\widetilde{\boldsymbol{M}} = \boldsymbol{C}'^{-1}\boldsymbol{M}\boldsymbol{C}^{-1}$, and $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{C}\boldsymbol{\Sigma}_*\boldsymbol{C}'$. Then

$$\begin{aligned}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu})'{\boldsymbol{\Sigma}_*}^{-1}\boldsymbol{M}^{-1}{\boldsymbol{\Sigma}_*}^{-1}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\nu}) &= (\boldsymbol{U} - \widetilde{\boldsymbol{\nu}})'\boldsymbol{C}'^{-1}\boldsymbol{\Sigma}_*^{-1}\boldsymbol{M}^{-1}\boldsymbol{\Sigma}_*^{-1}\boldsymbol{C}^{-1}(\boldsymbol{U} - \widetilde{\boldsymbol{\nu}})\\ &= (\boldsymbol{U} - \widetilde{\boldsymbol{\nu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{M}}^{-1}\widetilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{U} - \widetilde{\boldsymbol{\nu}}).\end{aligned}$$

Further, $\boldsymbol{m}_1 - \boldsymbol{\mu} = \boldsymbol{C}^{-1}(\boldsymbol{U} - \widetilde{\boldsymbol{\mu}})$, and

$$\boldsymbol{m}_2 - \boldsymbol{\mu} = \boldsymbol{C}^{-1}\left(\Big(\boldsymbol{I}_{p+q} - \frac{r\big((\boldsymbol{U} - \widetilde{\boldsymbol{\nu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{M}}^{-1}\widetilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{U} - \widetilde{\boldsymbol{\nu}})\big)}{(\boldsymbol{U} - \widetilde{\boldsymbol{\nu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{M}}^{-1}\widetilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{U} - \widetilde{\boldsymbol{\nu}})}\widetilde{\boldsymbol{M}}^{-1}\widetilde{\boldsymbol{\Sigma}}^{-1}\Big)(\boldsymbol{U} - \widetilde{\boldsymbol{\nu}}) + \widetilde{\boldsymbol{\nu}} - \widetilde{\boldsymbol{\mu}}\right).$$

Since

$$\boldsymbol{C}^{-1} = \begin{pmatrix}\boldsymbol{I}_p & \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ \boldsymbol{0} & \boldsymbol{I}_q\end{pmatrix}$$

then, by applying the assumption $\boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1} = -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, it is straightforward to verify that

$$\widetilde{\boldsymbol{M}} = \begin{pmatrix} \boldsymbol{M}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_{22} - \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{M}_{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{M}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_{22\cdot 1} \end{pmatrix}.$$

Similarly, we obtain

$$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{C}\boldsymbol{\Sigma}_*\boldsymbol{C}' = \begin{pmatrix} (\frac{1}{N} + \frac{q\bar{\tau}}{n(n-q-2)})\boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{N}\boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Let $\widetilde{\boldsymbol{m}}_1 = \boldsymbol{U}$ and

$$\widetilde{\boldsymbol{m}}_2 = \left(\boldsymbol{I}_{p+q} - \frac{r\big((\boldsymbol{U}-\widetilde{\boldsymbol{\nu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{M}}^{-1}\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{C}'^{-1}(\boldsymbol{U}-\widetilde{\boldsymbol{\nu}})\big)}{(\boldsymbol{U}-\widetilde{\boldsymbol{\nu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{M}}^{-1}\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{C}'^{-1}(\boldsymbol{U}-\widetilde{\boldsymbol{\nu}})} \widetilde{\boldsymbol{M}}^{-1}\widetilde{\boldsymbol{\Sigma}}^{-1} \right) (\boldsymbol{U}-\widetilde{\boldsymbol{\nu}}) + \widetilde{\boldsymbol{\nu}}.$$

For $i = 1, 2$, the loss $L(\boldsymbol{\mu}, \boldsymbol{m}_i) = (\boldsymbol{m}_i - \boldsymbol{\mu})'\boldsymbol{M}(\boldsymbol{m}_i - \boldsymbol{\mu})$ is transformed to

$$\begin{aligned} \widetilde{L}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{m}}_i) &= (\boldsymbol{C}'^{-1}(\widetilde{\boldsymbol{m}}_i - \widetilde{\boldsymbol{\mu}}))'\boldsymbol{M}(\boldsymbol{C}'^{-1}(\widetilde{\boldsymbol{m}}_i - \widetilde{\boldsymbol{\mu}})) \\ &= (\widetilde{\boldsymbol{m}}_i - \widetilde{\boldsymbol{\mu}})'\boldsymbol{C}^{-1}\boldsymbol{M}\boldsymbol{C}'^{-1}(\widetilde{\boldsymbol{m}}_i - \widetilde{\boldsymbol{\mu}}) \\ &= (\widetilde{\boldsymbol{m}}_i - \widetilde{\boldsymbol{\mu}})'\widetilde{\boldsymbol{M}}(\widetilde{\boldsymbol{m}}_i - \widetilde{\boldsymbol{\mu}}); \end{aligned}$$

therefore $\Delta R(\boldsymbol{\mu}) = E_{\widehat{\boldsymbol{\mu}}}\big[L(\boldsymbol{\mu}, \boldsymbol{m}_i) - L(\boldsymbol{\mu}, \boldsymbol{m}_i)\big] = E_{\boldsymbol{U}}\big[\widetilde{L}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{m}}_i) - \widetilde{L}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{m}}_i)\big]$. Noting that $\widetilde{\boldsymbol{M}}$ and $\widetilde{\boldsymbol{\Sigma}}$ both are block-diagonal matrices, we apply Theorem 4.2 to $\widetilde{L}$ to deduce that $\Delta R(\boldsymbol{\mu}) \geq 0$.

Finally, for the case in which $\boldsymbol{M} = \boldsymbol{\Sigma}^{-1}$, it is well-known that the condition $\boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1} = -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ is valid. $\square$

We remark that, more general than $\boldsymbol{M} = \boldsymbol{\Sigma}^{-1}$, the condition $\boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1} = -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ also holds if $\boldsymbol{M} = (c\boldsymbol{\Sigma} + \boldsymbol{\Lambda})^{-1}$ where $c > 0$, $\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$, and $\boldsymbol{\Lambda}_{11}$ is an arbitrary positive definite matrix.

5 Shrinkage to a common value

The results above show that under certain conditions on p, q, n, and N, the various shrinkage estimators $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ provide lower risk than the maximum likelihood estimator $\widehat{\boldsymbol{\mu}}$. However, as in the classical case of complete samples, usage of these shrinkage estimators requires prior specification of the vector $\boldsymbol{\nu}$. Although $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ provides lower risk than $\widehat{\boldsymbol{\mu}}$ irrespective of the true value of $\boldsymbol{\mu}$, the difference in risks between $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ and $\boldsymbol{\mu}$ is substantial as $\boldsymbol{\nu}$ draws closer to $\boldsymbol{\mu}$, and the difference is negligible if $\boldsymbol{\nu}$ is far from $\boldsymbol{\mu}$. In practical settings, it therefore is advantageous to choose $\boldsymbol{\nu}$ as close as possible to $\boldsymbol{\mu}$, a goal which is unlikely to be realistic given that $\boldsymbol{\mu}$ generally is unknown. To address this issue, Lindley and Smith [28] and Efron and Morris [19] developed shrinkage estimators in which each component of $\boldsymbol{\nu}$ has a common value computed from the sample mean. We now extend those results to the setting of monotone incomplete data.

We denote by $\boldsymbol{1}_k$ the k-dimensional column vector $(1, \ldots, 1)'$, with each entry equal to 1. In shrinking to a common mean, we assume that $\boldsymbol{\Sigma} = \boldsymbol{I}_{p+q}$ and adopt the standard squared-error loss function (1.2). Letting

$$\widehat{\nu} = \frac{1}{p+q}\boldsymbol{1}_{p+q}'\widehat{\boldsymbol{\mu}} \in \mathbb{R}, \tag{5.1}$$

we construct an estimator of James-Stein type,

$$\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c) = \left(1 - \frac{c}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2}\right)(\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}) + \widehat{\nu}\,\mathbf{1}_{p+q}. \tag{5.2}$$

We then have the following result.

Theorem 5.1. *Suppose that $p \ge 3$, $\boldsymbol{\Sigma} = \boldsymbol{I}_{p+q}$, the loss function is (1.2), and define*

$$c^* = \left(1 - \frac{1}{p+q}\right)\left(\frac{p}{n} + \frac{q}{N}\right) - \frac{2}{n}.$$

Then, for all $c \in (0, 2c^*)$, the shrinkage estimator $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ in (5.2) has smaller risk than $\widehat{\boldsymbol{\mu}}$.

Proof. As usual, we let $\Delta R(\boldsymbol{\mu}) = E[L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) - L\big(\boldsymbol{\mu}, \boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)\big)]$, the difference between the risks of $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$. Applying the parallelogram law (3.2), we obtain

$$\begin{aligned}\Delta R(\boldsymbol{\mu}) &= E\Big\{\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 - \Big\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu} - \frac{c}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2}(\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q})\Big\|^2\Big\} \\ &= E\Big\{\frac{c}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\mathbf{1}_{p+q}\|^2}\big[2(\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q})'(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) - c\big]\Big\}.\end{aligned}$$

We now apply Stein's formula (2.9) with $d \equiv p+q$, $\boldsymbol{Z} \equiv \widehat{\boldsymbol{\mu}}|Q$, $\boldsymbol{\theta} \equiv \boldsymbol{\mu}$, and $\boldsymbol{\Lambda} \equiv \mathrm{Cov}(\widehat{\boldsymbol{\mu}}|Q)$. Letting $\boldsymbol{z} = (z_1, \ldots, z_{p+q})'$ and $\bar{z} = \mathbf{1}'\boldsymbol{z}/(p+q)$, we define

$$\psi_i(\boldsymbol{z}) = \frac{z_i - \bar{z}}{\sum_{j=1}^{p+q}(z_j - \bar{z})^2},$$

$1 \le i \le p+q$, and then it is straightforward to verify that

$$\frac{\partial}{\partial z_i}\psi_i(\boldsymbol{z}) = \frac{\big(1 - (p+q)^{-1}\big)}{\|\boldsymbol{z} - \bar{z}\,\mathbf{1}_{p+q}\|^2} - \frac{2(z_i - \bar{z})^2}{\|\boldsymbol{z} - \bar{z}\,\mathbf{1}_{p+q}\|^4}.$$

By (2.9),

$$E_{\widehat{\boldsymbol{\mu}}|Q}\,\frac{(\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q})'(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2} = E_{\widehat{\boldsymbol{\mu}}|Q}\sum_{i=1}^{p+q}\left(\frac{1 - (p+q)^{-1}}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2} - \frac{2(\widehat{\mu}_i - \widehat{\nu})^2}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\right)\mathrm{Var}(\widehat{\mu}_i|Q),$$

where $\widehat{\mu}_1, \ldots, \widehat{\mu}_{p+q}$ are the components of $\widehat{\boldsymbol{\mu}}$. Substituting for $\mathrm{Var}(\widehat{\mu}_i|Q)$ from (2.8), we obtain

$$\begin{aligned}\Delta R(\boldsymbol{\mu}) &= E\Bigg\{\frac{2c(1+\bar{\tau}Q)}{n}\sum_{i=1}^{p}\left(\frac{1-(p+q)^{-1}}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2} - \frac{2(\widehat{\mu}_i - \widehat{\nu})^2}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\right) \\ &\qquad + \frac{2c}{N}\sum_{i=p+1}^{p+q}\left(\frac{1-(p+q)^{-1}}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2} - \frac{2(\widehat{\mu}_i - \widehat{\nu})^2}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\right) - \frac{c^2}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2}\Bigg\} \\ &= E\Bigg\{\frac{2c(1+\bar{\tau}Q)}{n}\left[\frac{\big(1-(p+q)^{-1}\big)\,p}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2} - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} - \widehat{\nu}\,\mathbf{1}_p\|^2}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\right] \\ &\qquad + \frac{2c}{N}\left[\frac{\big(1-(p+q)^{-1}\big)\,q}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2} - \frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} - \widehat{\nu}\,\mathbf{1}_q\|^2}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\right] - \frac{c^2}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2}\Bigg\}.\end{aligned}$$

Clearly, $1+\bar{\tau}Q \ge 1$; also, since $p \ge 3$ then $\big(1-(p+q)^{-1}\big)\,p \ge 2$. Therefore

$$\begin{aligned}
\frac{\big(1-(p+q)^{-1}\big)\,p}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}-\frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\widehat{\nu}\,\mathbf{1}_p\|^2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^4} &\ge \frac{2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}-\frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\widehat{\nu}\,\mathbf{1}_p\|^2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\\
&= \frac{2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\left(\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2-\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\widehat{\nu}\,\mathbf{1}_p\|^2\right)\\
&= \frac{2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\widehat{\nu}\,\mathbf{1}_q\|^2 \ge 0.
\end{aligned}$$

Therefore

$$\begin{aligned}
\Delta R(\boldsymbol{\mu}) &\ge E\Bigg\{\frac{2c}{n}\left[\frac{\big(1-(p+q)^{-1}\big)\,p}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}-\frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\widehat{\nu}\,\mathbf{1}_p\|^2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\right]\\
&\qquad +\frac{2c}{N}\left[\frac{\big(1-(p+q)^{-1}\big)\,q}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}-\frac{2\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\widehat{\nu}\,\mathbf{1}_q\|^2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^4}\right]-\frac{c^2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}\Bigg\}\\
&= E\Bigg\{2c\big(1-(p+q)^{-1}\big)\frac{pn^{-1}+qN^{-1}}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}\\
&\qquad -4\frac{n^{-1}\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\widehat{\nu}\,\mathbf{1}_p\|^2+N^{-1}\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\widehat{\nu}\,\mathbf{1}_q\|^2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^4}-\frac{c^2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}\Bigg\}.
\end{aligned}$$

Since $n \le N$ then

$$\begin{aligned}
n^{-1}\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\widehat{\nu}\,\mathbf{1}_p\|^2+N^{-1}\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\widehat{\nu}\,\mathbf{1}_q\|^2 &\le n^{-1}\left(\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}-\widehat{\nu}\,\mathbf{1}_p\|^2+\|\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}-\widehat{\nu}\,\mathbf{1}_q\|^2\right)\\
&= n^{-1}\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2.
\end{aligned}$$

Hence,

$$\begin{aligned}
\Delta R(\boldsymbol{\mu}) &\ge E\left\{2c\big(1-(p+q)^{-1}\big)\frac{pn^{-1}+qN^{-1}}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}-\frac{4\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}{n\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^4}-\frac{c^2}{\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^2}\right\}\\
&= c\left(2\big(1-(p+q)^{-1}\big)\left(\frac{p}{n}+\frac{q}{N}\right)-\frac{4}{n}-c\right)E\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^{-2}\\
&= c(2c^*-c)\,E\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^{-2} > 0,
\end{aligned}$$

for $0 < c < 2c^*$.

Finally, it remains to be verified that if $p \ge 3$ then (i) $c^* > 0$ and (ii) $E\|\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}\|^{-2} < \infty$. To prove (i), note that c^* is an increasing function of p or q; therefore, for $p \ge 3$ and $q \ge 1$,

$$c^* \ge \left(1-\frac{1}{4}\right)\left(\frac{3}{n}+\frac{1}{N}\right)-\frac{2}{n}=\frac{1}{4n}+\frac{3}{4N} > 0.$$

To prove (ii), we note, first, that

$$\widehat{\boldsymbol{\mu}}-\widehat{\nu}\,\mathbf{1}_{p+q}=\boldsymbol{J}\widehat{\boldsymbol{\mu}} \tag{5.3}$$

where $\boldsymbol{J}=\boldsymbol{I}_{p+q}-(p+q)^{-1}\,\mathbf{1}_{p+q}\,\mathbf{1}_{p+q}'$ is symmetric, idempotent, and of rank $p+q-1$. Then there exists a $(p+q)\times(p+q)$ orthogonal matrix $\boldsymbol{H}$ such that

$$\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}'=\begin{pmatrix}\boldsymbol{I}_{p+q-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}.$$

Let $\boldsymbol{U} = \boldsymbol{H}\widehat{\boldsymbol{\mu}}$; then we have

$$\begin{aligned} \|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2 &= \|\boldsymbol{J}\widehat{\boldsymbol{\mu}}\|^2 = \widehat{\boldsymbol{\mu}}'\boldsymbol{J}^2\widehat{\boldsymbol{\mu}} \\ &= \widehat{\boldsymbol{\mu}}'\boldsymbol{J}\widehat{\boldsymbol{\mu}} = \boldsymbol{U}'\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}'\boldsymbol{U} = \sum_{j=1}^{p+q-1} U_j^2, \end{aligned} \tag{5.4}$$

where $U_1, \ldots, U_{p+q}$ are the components of $\boldsymbol{U}$. Next, let $\boldsymbol{K} = (\boldsymbol{I}_{p+q-1} \vdots \mathbf{0})$ be the $(p+q-1) \times (p+q)$ matrix in which the first $p+q-1$ rows and columns consist of the identity matrix $\boldsymbol{I}_{p+q-1}$ and the last column consists entirely of zeros. Transforming $\boldsymbol{U}$ to $\widetilde{\boldsymbol{U}} = \boldsymbol{K}\boldsymbol{U}$, we have $\widetilde{\boldsymbol{U}} = (U_1, \ldots, U_{p+q-1})'$, and therefore $\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2 = \widetilde{\boldsymbol{U}}'\widetilde{\boldsymbol{U}}$.

By Corollary 2.2, $\widehat{\boldsymbol{\mu}}|Q \sim N_{p+q}(\boldsymbol{\mu}, \boldsymbol{C}_Q)$, where $\boldsymbol{C}_Q$ is given in (2.8). Therefore, conditional on Q, $\widetilde{\boldsymbol{U}} = \boldsymbol{K}\boldsymbol{U} = \boldsymbol{K}\boldsymbol{H}\widehat{\boldsymbol{\mu}} \sim N_{p+q-1}(\boldsymbol{K}\boldsymbol{H}\boldsymbol{\mu}, \boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'\boldsymbol{K}')$, and this normal distribution is nonsingular because $\boldsymbol{K}$ is of full rank and both $\boldsymbol{H}$ and $\boldsymbol{C}_Q$ are nonsingular. By Lemma 2.4,

$$E\big((\widetilde{\boldsymbol{U}}'\widetilde{\boldsymbol{U}})^{-1}|Q\big) \le \frac{1}{(p+q-3)\,\lambda_{(p+q-1)}(\boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'\boldsymbol{K}')},$$

where $\lambda_{(p+q-1)}(\boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'\boldsymbol{K}')$ is the smallest eigenvalue of $\boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'\boldsymbol{K}'$.

Since $\lambda_{(p+q-1)}(\boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'\boldsymbol{K}')$ depends on Q, we need to show that

$$E_Q\left(1/\lambda_{(p+q-1)}(\boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'\boldsymbol{K}')\right) < \infty.$$

Noting that $\boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'\boldsymbol{K}'$ is the upper $(p+q-1)\times(p+q-1)$ principal submatrix of $\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'$, we apply Cauchy's Interlacing Theorem (see Bhatia [10], Corollary III.1.5, p. 59) to deduce that $\lambda_{(p+q-1)}(\boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'\boldsymbol{K}') \ge \lambda_{(p+q)}(\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}')$, the smallest eigenvalue of $\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'$. Since $\boldsymbol{H}$ is orthogonal then $\lambda_{(p+q)}(\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}') = \lambda_{(p+q)}(\boldsymbol{C}_Q)$; therefore

$$\lambda_{(p+q-1)}(\boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_Q\boldsymbol{H}'\boldsymbol{K}') \ge \lambda_{(p+q)}(\boldsymbol{C}_Q) = \frac{1}{N}.$$

Hence

$$E\big((\widetilde{\boldsymbol{U}}'\widetilde{\boldsymbol{U}})^{-1}|Q\big) \le \frac{1}{(p+q-3)\lambda_{(p+q)}(\boldsymbol{C}_Q)} \le \frac{N}{p+q-3},$$

and so we obtain the same bound on the unconditional expectation,

$$E\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^{-2} = E\big((\widetilde{\boldsymbol{U}}'\widetilde{\boldsymbol{U}})^{-1}\big) \le \frac{N}{p+q-3}. \tag{5.5}$$

The proof of the theorem now is complete. $\square$

We turn next to the positive-part estimator for shrinking to a common mean. Here, we use the notation t_+, $t \in \mathbb{R}$, defined in (3.13). Define

$$\boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c) = \left(1 - \frac{c}{\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2}\right)_+ (\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}) + \widehat{\nu}\,\mathbf{1}_{p+q},$$

$c \ge 0$, where $\widehat{\nu}$ again is given in (5.1). We then have the following result.

Theorem 5.2. *Suppose that $p \geq 3$, $\boldsymbol{\Sigma} = \boldsymbol{I}_{p+q}$, $\mathbf{1}'\boldsymbol{\mu} \neq 0$, and define $\bar{\mu} = (\mathbf{1}_{p+q}'\boldsymbol{\mu})/(p+q)$. Further, let (1.2) be the loss function.*

(i) Suppose that $\boldsymbol{\mu}_0 = a\,\mathbf{1}_{p+q}$ for some constant $a \in \mathbb{R}$. Then there exists an open neighborhood $\mathcal{N}$ of $\boldsymbol{\mu}_0$ such that, for any $c > 0$, $\boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c)$ has smaller risk than $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ for all $\boldsymbol{\mu} \in \mathcal{N}$.

(ii) Suppose that $\boldsymbol{\mu} \neq a\,\mathbf{1}_{p+q}$ for any $a \in \mathbb{R}$. Then there exists a constant $c^* > E\,\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2$ such that the positive-part estimator $\boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c)$ has smaller risk than $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ for all $c > c^*$. In particular,

$$E\,\|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2 = \|\boldsymbol{\mu} - \bar{\mu}\,\mathbf{1}_{p+q}\|^2 + \frac{p+q-1}{p+q}\left[\frac{p}{n\eta_*} + \frac{q}{N}\right]. \tag{5.6}$$

Proof. Let $g(t) = 1 - ct^{-2}$, $t > 0$. Also, let $\boldsymbol{U} = \widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}$. Then

$$\begin{aligned} L\big(\boldsymbol{\mu}, \boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)\big) &- L\big(\boldsymbol{\mu}, \boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c)\big) \\ &= \big\|\widehat{\nu}\,\mathbf{1}_{p+q} - \boldsymbol{\mu} + g(\|\boldsymbol{U}\|)\boldsymbol{U}\big\|^2 - \big\|\widehat{\nu}\,\mathbf{1}_{p+q} - \boldsymbol{\mu} + g^+(\|\boldsymbol{U}\|)\boldsymbol{U}\big\|^2 \\ &= [g(\|\boldsymbol{U}\|)]^2\|\boldsymbol{U}\|^2 + 2g(\|\boldsymbol{U}\|)\boldsymbol{U}'(\widehat{\nu}\,\mathbf{1}_{p+q} - \boldsymbol{\mu}) \\ &\qquad - [g^+(\|\boldsymbol{U}\|)]^2\|\boldsymbol{U}\|^2 - 2g^+(\|\boldsymbol{U}\|)\boldsymbol{U}'(\widehat{\nu}\,\mathbf{1}_{p+q} - \boldsymbol{\mu}). \end{aligned} \tag{5.7}$$

Since $\widehat{\nu} = (p+q)^{-1}\,\mathbf{1}_{p+q}'\widehat{\boldsymbol{\mu}}$ then

$$\begin{aligned} \boldsymbol{U}'\,\mathbf{1}_{p+q} &= (\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q})'\mathbf{1}_{p+q} \\ &= \widehat{\boldsymbol{\mu}}'\,\mathbf{1}_{p+q} - \widehat{\nu}\,\mathbf{1}_{p+q}'\,\mathbf{1}_{p+q} = \mathbf{1}_{p+q}'\widehat{\boldsymbol{\mu}} - (p+q)\widehat{\nu} = 0, \end{aligned}$$

and therefore $\boldsymbol{U}'(\widehat{\nu}\,\mathbf{1}_{p+q} - \boldsymbol{\mu}) = -\boldsymbol{U}'\boldsymbol{\mu}$. It now follows from (5.7) that

$$\Delta R(\boldsymbol{\mu}) = E[L\big(\boldsymbol{\mu}, \boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)\big) - L\big(\boldsymbol{\mu}, \boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c)\big)],$$

the difference in risks between $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ and $\boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c)$, is given by

$$\Delta R(\boldsymbol{\mu}) = E\left[\big([g(\|\boldsymbol{U}\|)]^2 - [g^+(\|\boldsymbol{U}\|)]^2\big)\,\|\boldsymbol{U}\|^2 + 2\big(g^+(\|\boldsymbol{U}\|) - g(\|\boldsymbol{U}\|)\big)\boldsymbol{\mu}'\boldsymbol{U}\right].$$

We can now establish part (i) of the theorem. Suppose that $\boldsymbol{J}\boldsymbol{\mu}_0 = 0$; then $\boldsymbol{\mu}_0'\boldsymbol{U} = \boldsymbol{\mu}_0'\boldsymbol{J}\widehat{\boldsymbol{\mu}} = (\boldsymbol{J}\boldsymbol{\mu}_0)'\widehat{\boldsymbol{\mu}} = 0$, therefore

$$\Delta R(\boldsymbol{\mu}_0) = E\,\big([g(\|\boldsymbol{U}\|)]^2 - [g^+(\|\boldsymbol{U}\|)]^2\big)\,\|\boldsymbol{U}\|^2.$$

Denote by $\mathcal{I}_{(a,b)}$ the indicator function of the interval (a,b). Then, by the definition of g^+,

$$[g(t)]^2 - [g^+(t)]^2 = \begin{cases} 0, & t \geq c^{1/2} \\ [g(t)]^2, & t < c^{1/2} \end{cases} \equiv [g(t)]^2\mathcal{I}_{(0,c^{1/2})}(t);$$

therefore $[g(t)]^2 - [g^+(t)]^2 \geq 0$ for all t, so $\Delta R(\boldsymbol{\mu}_0) \geq 0$. Note that $[g(t)]^2 > 0$ on the interval $(0, c^{1/2})$. Moreover, by (5.4), the distribution of $\|\boldsymbol{U}\|^2$, and hence of $\|\boldsymbol{U}\|$, is non-degenerate on the interval $0 < \|\boldsymbol{U}\| < c^{1/2}$. Therefore $\Delta R(\boldsymbol{\mu}_0) > 0$, and it follows by continuity of ΔR as a function on $\mathbb{R}^{p+q}$ that there exists an open neighborhood $\mathcal{N} \subset \mathbb{R}^{p+q}$ of $\boldsymbol{\mu}_0$ such that $\Delta R(\boldsymbol{\mu}) > 0$ on $\mathcal{N}$. To conclude the proof of (i), we observe that $\boldsymbol{J}\boldsymbol{\mu}_0 = 0$ if and only if

$$\begin{aligned} 0 &= \big(\boldsymbol{I}_{p+q} - (p+q)^{-1}\,\mathbf{1}_{p+q}\,\mathbf{1}_{p+q}'\big)\boldsymbol{\mu}_0 \\ &= \boldsymbol{\mu}_0 - (p+q)^{-1}\,\mathbf{1}_{p+q}(\,\mathbf{1}_{p+q}'\boldsymbol{\mu}_0) = \boldsymbol{\mu}_0 - a\,\mathbf{1}_{p+q}, \end{aligned}$$

where $a = (p+q)^{-1}(\,\mathbf{1}'_{p+q}\boldsymbol{\mu}_0)$.

We now prove (ii). Observe that $g(t) - g^{+}(t) \equiv -g(t)\mathcal{I}_{(0,c^{1/2})}(t)$, and therefore

$$\Delta R(\boldsymbol{\mu}) = E\left[[g(\|\boldsymbol{U}\|)]^2\|\boldsymbol{U}\|^2 - 2g(\|\boldsymbol{U}\|)\boldsymbol{\mu}'\boldsymbol{U}\right]\mathcal{I}_{(0,c^{1/2})}(\|\boldsymbol{U}\|).$$

Let $T = \|\boldsymbol{U}\|$; then, by the Cauchy-Schwartz inequality,

$$\begin{aligned} E\,[g(\|\boldsymbol{U}\|)]^2\,\|\boldsymbol{U}\|^2\,\mathcal{I}_{(0,c^{1/2})}(\|\boldsymbol{U}\|) &\equiv E\,[g(T)]^2\,T^2\,\mathcal{I}_{(0,c^{1/2})}(T) \\ &\geq \Big(E\,T\,g(T)\,\mathcal{I}_{(0,c^{1/2})}(T)\Big)^2 \end{aligned}$$

and also $\boldsymbol{\mu}'\boldsymbol{U} \leq \|\boldsymbol{\mu}\|\,T$, so that

$$\big|E\,g(\|\boldsymbol{U}\|)\,\boldsymbol{\mu}'\boldsymbol{U}\,\mathcal{I}_{(0,c^{1/2})}(\|\boldsymbol{U}\|)\big| \leq \|\boldsymbol{\mu}\|\,E\,T\,|g(T)|\,\mathcal{I}_{(0,c^{1/2})}(T).$$

Therefore

$$\Delta R(\boldsymbol{\mu}) \geq \Big(E\,T\,|g(T)|\,\mathcal{I}_{(0,c^{1/2})}(T)\Big)\Big(E\,T\,|g(T)|\,\mathcal{I}_{(0,c^{1/2})}(T) - 2\|\boldsymbol{\mu}\|\Big)\,. \tag{5.8}$$

Noting that $t\,|g(t)| = t^{-1}\,|c - t^2|$, we obtain

$$E\,T\,|g(T)|\,\mathcal{I}_{(0,c^{1/2})}(T) = E\,(c - T^2)\,T^{-1}\,\mathcal{I}_{(0,c^{1/2})}(T).$$

Since the functions $c - t^2$ and $t^{-1}\mathcal{I}_{(0,c^{1/2})}(t)$, $t > 0$, both are monotonic decreasing then, by Chebyshev's other inequality [21],

$$E\,(c - T^2)\,T^{-1}\,\mathcal{I}_{(0,c^{1/2})}(T) \geq E\,(c - T^2)\cdot E\,T^{-1}\,\mathcal{I}_{(0,c^{1/2})}(T).$$

Since $t^{-1} > c^{-1/2}$ whenever $0 < t < c^{1/2}$ then

$$\begin{aligned} E\,T^{-1}\,\mathcal{I}_{(0,c^{1/2})}(T) &\geq c^{-1/2}E\,\mathcal{I}_{(0,c^{1/2})}(T) \\ &= c^{-1/2}P(T < c^{1/2}) \\ &\geq \max\big\{c^{-1/2}\big(1 - c^{-1}\,E(T^2)\big), 0\big\}, \end{aligned}$$

where the latter bound holds by virtue of Markov's inequality. Therefore

$$\begin{aligned} E\,T\,|g(T)|\,\mathcal{I}_{(0,c^{1/2})}(T) &\geq (c - E(T^2))\cdot\max\big\{c^{-1/2}\big(1 - c^{-1}\,E(T^2)\big), 0\big\} \\ &= c^{1/2}(1 - c^{-1}E(T^2))\cdot\max\big\{\big(1 - c^{-1}\,E(T^2)\big), 0\big\}. \end{aligned}$$

Define the function

$$h(t) = t^{1/2}(1 - t^{-1}E(T^2))\cdot\max\big\{\big(1 - t^{-1}\,E(T^2)\big), 0\big\},$$

$t > 0$; then h is strictly increasing on the interval $(E(T^2), \infty)$ and so the equation $h(t) = 2\|\boldsymbol{\mu}\|$ has a unique solution $c^* \in (E(T^2), \infty)$. Therefore, by (5.8), $\Delta R(\boldsymbol{\mu}) > 0$ for all $c \in (c^*, \infty)$.

Finally, it remains to establish (5.6), the explicit formula for $E(\|\boldsymbol{U}\|^2)$. We note first that

$$\|\boldsymbol{U}\|^2 = \|\widehat{\boldsymbol{\mu}} - \widehat{\nu}\,\mathbf{1}_{p+q}\|^2 = \|\widehat{\boldsymbol{\mu}}\|^2 - (p+q)^{-1}(\,\mathbf{1}'_{p+q}\widehat{\boldsymbol{\mu}})^2.$$

Applying Corollary 2.2, which provides that $\widehat{\boldsymbol{\mu}}|Q \sim N_{p+q}(\boldsymbol{\mu}, \boldsymbol{C}_Q)$, we obtain

$$E(\|\widehat{\boldsymbol{\mu}}\|^2|Q) = \boldsymbol{\mu}'\boldsymbol{\mu} + \mathrm{tr}(\boldsymbol{C}_Q) = \|\boldsymbol{\mu}\|^2 + \frac{p}{n}(1+\bar{\tau}Q) + \frac{q}{N}.$$

Also, $\mathbf{1}'_{p+q}\widehat{\boldsymbol{\mu}}|Q \sim N\left(\mathbf{1}'_{p+q}\boldsymbol{\mu}, \frac{p}{n}(1+\bar{\tau}Q) + \frac{q}{N}\right)$ and hence

$$E(\mathbf{1}'_{p+q}\widehat{\boldsymbol{\mu}})^2 = (\mathbf{1}'_{p+q}\boldsymbol{\mu})^2 + \frac{p}{n}(1+\bar{\tau}Q) + \frac{q}{N}.$$

Therefore

$$\begin{aligned} E(\|\boldsymbol{U}\|^2|Q) &= \|\boldsymbol{\mu}\|^2 + \frac{p}{n}(1+\bar{\tau}Q) + \frac{q}{N} - (p+q)^{-1}\left[(\mathbf{1}'_{p+q}\boldsymbol{\mu})^2 + \frac{p}{n}(1+\bar{\tau}Q) + \frac{q}{N}\right] \\ &= \|\boldsymbol{\mu}\|^2 - (p+q)^{-1}(\mathbf{1}'_{p+q}\boldsymbol{\mu})^2 + \frac{p+q-1}{p+q}\left[\frac{p}{n}(1+\bar{\tau}Q) + \frac{q}{N}\right]. \end{aligned}$$

On taking expectations with respect to Q, we obtain

$$\begin{aligned} E(\|\boldsymbol{U}\|^2) &= \|\boldsymbol{\mu}\|^2 - \frac{1}{p+q}(\mathbf{1}'_{p+q}\boldsymbol{\mu})^2 + \frac{p+q-1}{p+q}\left[\frac{p}{n}\big(1+\bar{\tau}E(Q)\big) + \frac{q}{N}\right] \\ &= \|\boldsymbol{\mu}\|^2 - \frac{1}{p+q}(\mathbf{1}'_{p+q}\boldsymbol{\mu})^2 + \frac{p+q-1}{p+q}\left[\frac{p}{n}\big(1+\frac{q\bar{\tau}}{n-q-2}\big) + \frac{q}{N}\right]; \end{aligned}$$

and the conclusion follows from the elementary identity,

$$\|\boldsymbol{\mu}\|^2 - (p+q)^{-1}(\mathbf{1}'_{p+q}\boldsymbol{\mu})^2 = \|\boldsymbol{\mu} - \bar{\mu}\,\mathbf{1}_{p+q}\|^2,$$

where, as defined earlier, $\bar{\mu} = (\mathbf{1}'_{p+q}\boldsymbol{\mu})/(p+q)$. $\square$

We remark that it seems difficult to analyze $\Delta R(\boldsymbol{\mu})$ for small values of $c > 0$, and this leads to the above result in which we can establish positivity of $\Delta R(\boldsymbol{\mu})$ only for sufficiently large c. Motivated by this difficulty, we consider a modification to the estimator (5.2) that reflects the monotone nature of the data. If the investigator uses a two-part common-value shrinkage estimator then, as the following result shows, the natural relationship between the risk of $\widehat{\boldsymbol{\mu}}$ and the unrestricted and positive-part estimators returns.

As before, let $\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}$ be the vector of the first p components of $\widehat{\boldsymbol{\mu}}$, and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$ be the vector of the last q components of $\widehat{\boldsymbol{\mu}}$. Further, let $\widehat{\nu}_1 = p^{-1}\,\mathbf{1}'_p\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}$, $\widehat{\nu}_2 = q^{-1}\,\mathbf{1}'_q\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}}$, and

$$\widehat{\boldsymbol{\nu}} = \begin{pmatrix} \widehat{\nu}_1\,\mathbf{1}_p \\ \widehat{\nu}_2\,\mathbf{1}_q \end{pmatrix}. \tag{5.9}$$

For $c > 0$, define the shrinkage estimators

$$\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c) = \left(1 - \frac{c}{\|\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\nu}}\|^2}\right)(\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\nu}}) + \widehat{\boldsymbol{\nu}}$$

and

$$\boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c) = \left(1 - \frac{c}{\|\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\nu}}\|^2}\right)_+(\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\nu}}) + \widehat{\boldsymbol{\nu}}$$

We now have the following result.

Theorem 5.3. *For $p \ge 3$ and $q \ge 2$, and with the loss function (1.2), we have*

$$R(\boldsymbol{\mu}; \boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c)) < R(\boldsymbol{\mu}; \boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)) < R(\boldsymbol{\mu}; \widehat{\boldsymbol{\mu}})$$

for all $c \in (0, 2c^{**})$ where

$$c^{**} = \frac{p-3}{n} + \frac{q-1}{N}.$$

In particular, $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ are inadmissible.

Proof. Let $\boldsymbol{H}_1$ be a $p \times p$ orthogonal matrix whose first row is $p^{-1/2}\,\mathbf{1}'_p$ and let $\boldsymbol{H}_2$ be a $q \times q$ orthogonal matrix with last row $q^{-1/2}\,\mathbf{1}'_q$. Define

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2 \end{pmatrix},$$

set $\boldsymbol{U} = \boldsymbol{H}\widehat{\boldsymbol{\mu}}$, and denote by $U_1, \ldots, U_{p+q}$ the components of $\boldsymbol{U}$. Introduce the notation $\boldsymbol{U}_{2,p+q-1} = (U_2, \ldots, U_{p+q-1})'$, so that $\boldsymbol{U} = (U_1, \boldsymbol{U}'_{2,p+q-1}, U_{p+q})'$.

Since $\boldsymbol{H}_1$ has first row $p^{-1/2}\,\mathbf{1}'_p$ then $U_1 = p^{-1/2}\,\mathbf{1}'_p\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} = p^{1/2}\widehat{\nu}_1$; similarly, $U_{p+q} = q^{-1/2}\,\mathbf{1}'_q\widehat{\boldsymbol{\mu}}_{\boldsymbol{Y}} = q^{1/2}\widehat{\nu}_q$. Since $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are orthogonal with their stated first and last rows, respectively, it follows from the definition of $\widehat{\boldsymbol{\nu}}$ in (5.9) that

$$\begin{aligned} \boldsymbol{H}\widehat{\boldsymbol{\nu}} &= \begin{pmatrix} \boldsymbol{H}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2 \end{pmatrix} \begin{pmatrix} \widehat{\nu}_1\,\mathbf{1}_p \\ \widehat{\nu}_2\,\mathbf{1}_q \end{pmatrix} = \begin{pmatrix} \widehat{\nu}_1\boldsymbol{H}_1\,\mathbf{1}_p \\ \widehat{\nu}_q\boldsymbol{H}_2\,\mathbf{1}_q \end{pmatrix} \\ &= (p^{1/2}\widehat{\nu}_1, 0, \ldots, 0, q^{1/2}\widehat{\nu}_2)' = (U_1, 0, \ldots, 0, U_{p+q})'; \end{aligned} \tag{5.10}$$

therefore $\boldsymbol{H}(\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\nu}}) = \boldsymbol{U} - (U_1, 0, \ldots, 0, U_{p+q})' = (0, \boldsymbol{U}_{2,p+q-1}, 0)'$. Since $\boldsymbol{H}$ is orthogonal then

$$\|\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\nu}}\|^2 = \|\boldsymbol{H}(\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\nu}})\|^2 = \|\boldsymbol{U}_{2,p+q-1}\|^2. \tag{5.11}$$

Let $\widetilde{\boldsymbol{\mu}} = \boldsymbol{H}\boldsymbol{\mu}$; similar to $\boldsymbol{U}_{2,p+q-1}$, define $\widetilde{\boldsymbol{\mu}}_{2,p+q-1} = (\widetilde{\mu}_2, \ldots, \widetilde{\mu}_{p+q-1})'$, and then we can express $\widetilde{\boldsymbol{\mu}}$ in the form $\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_1, \widetilde{\boldsymbol{\mu}}'_{2,p+q-1}, \widetilde{\mu}_{p+q})'$. Since $\boldsymbol{U} = \boldsymbol{H}\widehat{\boldsymbol{\mu}}$ then, by the stochastic representation (2.5),

$$\boldsymbol{U} \stackrel{\mathcal{L}}{=} \widetilde{\boldsymbol{\mu}} + \widetilde{\boldsymbol{V}}_1 + \sqrt{\frac{\bar{\tau}Q}{n}} \begin{pmatrix} \widetilde{\boldsymbol{V}}_2 \\ \mathbf{0} \end{pmatrix},$$

where $\widetilde{\boldsymbol{V}}_1 = \boldsymbol{H}\boldsymbol{V}_1 \sim N_{p+q}(\mathbf{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = \begin{pmatrix} \frac{1}{n}\boldsymbol{I}_p & \mathbf{0} \\ \mathbf{0} & \frac{1}{N}\boldsymbol{I}_q \end{pmatrix}$; $\widetilde{\boldsymbol{V}}_2 = \boldsymbol{H}_1\boldsymbol{V}_2 \sim N_p(\mathbf{0}, \boldsymbol{I}_p)$; and $\widetilde{\boldsymbol{V}}_1$ and $\widetilde{\boldsymbol{V}}_2$ are independent because $\boldsymbol{H}$ is orthogonal. Therefore, the marginal distribution of $\boldsymbol{U}_{2,p+q-1}$ is given by

$$\boldsymbol{U}_{2,p+q-1} \stackrel{\mathcal{L}}{=} \widetilde{\boldsymbol{\mu}}_{2,p+q-1} + \widetilde{\boldsymbol{\mathcal{V}}}_1 + \sqrt{\frac{\bar{\tau}Q}{n}} \begin{pmatrix} \widetilde{\boldsymbol{\mathcal{V}}}_2 \\ \mathbf{0} \end{pmatrix} \tag{5.12}$$

where $\widetilde{\boldsymbol{\mathcal{V}}}_1 \sim N_{p+q-2}(\mathbf{0}, \widetilde{\boldsymbol{\Omega}})$, $\widetilde{\boldsymbol{\Omega}} = \begin{pmatrix} \frac{1}{n}\boldsymbol{I}_{p-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{N}\boldsymbol{I}_{q-1} \end{pmatrix}$; $\widetilde{\boldsymbol{\mathcal{V}}}_2 \sim N_{p-1}(\mathbf{0}, \boldsymbol{I}_{p-1})$; and $\widetilde{\boldsymbol{\mathcal{V}}}_1$, $\widetilde{\boldsymbol{\mathcal{V}}}_2$, and Q are mutually independent. We see now that the distribution of $\boldsymbol{U}_{2,p+q-1}$ is obtained from the distribution of $\boldsymbol{U}$ by replacing p and q by $p-1$ and $q-1$, respectively.

Again by the orthogonality of $\boldsymbol{H}$,

$$\begin{aligned}\|\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c)-\boldsymbol{\mu}\|^2 &= \|\boldsymbol{H}\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c)-\boldsymbol{H}\boldsymbol{\mu}\|^2\\ &= \Big\|\boldsymbol{H}\widehat{\boldsymbol{\nu}}+\Big(1-\frac{c}{\|\widehat{\boldsymbol{\mu}}-\widehat{\boldsymbol{\nu}}\|^2}\Big)(\boldsymbol{H}\widehat{\boldsymbol{\mu}}-\boldsymbol{H}\widehat{\boldsymbol{\nu}})-\widetilde{\boldsymbol{\mu}}\Big\|^2.\end{aligned}$$

Applying (5.10) and (5.11), we obtain

$$\begin{aligned}&\|\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c)-\boldsymbol{\mu}\|^2\\ &\quad= (U_1-\widetilde{\mu}_1)^2+\Big\|\Big(1-\frac{c}{\|\boldsymbol{U}_{2,p+q-1}\|^2}\Big)\boldsymbol{U}_{2,p+q-1}-\widetilde{\boldsymbol{\mu}}_{2,p+q-1}\Big\|^2+(U_{p+q}-\widetilde{\mu}_{p+q})^2. \quad (5.13)\end{aligned}$$

Similarly,

$$\begin{aligned}&\|\boldsymbol{m}^+(\widehat{\boldsymbol{\mu}},c)-\boldsymbol{\mu}\|^2\\ &\quad= (U_1-\widetilde{\mu}_1)^2+\Big\|\Big(1-\frac{c}{\|\boldsymbol{U}_{2,p+q-1}\|^2}\Big)_+\boldsymbol{U}_{2,p+q-1}-\widetilde{\boldsymbol{\mu}}_{2,p+q-1}\|^2+(U_{p+q}-\widetilde{\mu}_{p+q})^2 \quad (5.14)\end{aligned}$$

and

$$\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\|^2 = (U_1-\widetilde{\mu}_1)^2+\|\boldsymbol{U}_{2,p+q-1}-\widetilde{\boldsymbol{\mu}}_{2,p+q-1}\|^2+(U_{p+q}-\widetilde{\mu}_{p+q})^2. \quad (5.15)$$

Subtracting (5.13) from (5.15) and taking expectations, we obtain

$$\begin{aligned}&R(\boldsymbol{\mu},\widehat{\boldsymbol{\mu}})-R(\boldsymbol{\mu},\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c))\\ &\qquad= R(\widetilde{\boldsymbol{\mu}}_{2,p+q-1},\boldsymbol{U}_{2,p+q-1})-R\Big(\widetilde{\boldsymbol{\mu}}_{2,p+q-1},(1-\frac{c}{\|\boldsymbol{U}_{2,p+q-1}\|^2})\boldsymbol{U}_{2,p+q-1}\Big).\end{aligned}$$

Noting that $\boldsymbol{U}_{2,p+q-1}$ has the stochastic representation (5.12), we now apply Theorem 3.1 with p and q replaced by $p-1$ and $q-1$, respectively, and deduce that $R(\boldsymbol{\mu},\widehat{\boldsymbol{\mu}})-R(\boldsymbol{\mu},\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c))>0$ for $0<c<2c^{**}$.

Similarly, subtracting (5.14) from (5.15) and taking expectations, we obtain

$$\begin{aligned}R(\boldsymbol{\mu},\boldsymbol{m}(\widehat{\boldsymbol{\mu}},c))-R(\boldsymbol{\mu},\boldsymbol{m}^+(\widehat{\boldsymbol{\mu}},c)) &= R\Big(\widetilde{\boldsymbol{\mu}}_{2,p+q-1},(1-\frac{c}{\|\boldsymbol{U}_{2,p+q-1}\|^2})\boldsymbol{U}_{2,p+q-1}\Big)\\ &\qquad - R\Big(\widetilde{\boldsymbol{\mu}}_{2,p+q-1},(1-\frac{c}{\|\boldsymbol{U}_{2,p+q-1}\|^2})_+\boldsymbol{U}_{2,p+q-1}\Big),\end{aligned}$$

and then the conclusion follows from (5.12) and Theorem 3.2 for the positive-part shrinkage estimator. □

6 Concluding Remarks

The results derived in this paper raise many issues, and we now present some of them as avenues for further research.

It is natural to ask for extensions of all the foregoing results to general k-step monotone incomplete data, to the incomplete data patterns considered by Eaton and Kariya [18], and even to arbitrary incomplete patterns. In all of these problems, there remains the fundamental impediment that the exact distribution of $\widehat{\boldsymbol{\mu}}$, the maximum likelihood estimator of $\boldsymbol{\mu}$, is still unknown. Indeed, in the case of non-monotone incomplete data patterns, explicit expressions for $\widehat{\boldsymbol{\mu}}$ are generally unavailable, so that the derivation of its exact distribution, and hence explicit formulas for lower bounds on the risk of $\widehat{\boldsymbol{\mu}}$, appears to require techniques far beyond those available today. In particular, it seems a formidable problem to study these problems when the EM, or other computational algorithms, are necessary to calculate $\widehat{\boldsymbol{\mu}}$. Nevertheless, we conjecture that, for any monotone incomplete data pattern and for sufficiently large p and n, the Stein phenomenon holds for the estimator $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ in (1.3).

We also find it to be remarkable that the shrinkage estimators have forms entirely reminiscent of the James-Stein type. This suggests an almost universal nature to these estimators in the presence of monotone incomplete samples. Even more, we raise the question of whether estimators of that type will prove to have lower risk than $\widehat{\mu}$ even in the case of non-monotone incomplete normal samples.

It also remains an open problem in Theorem 3.1 to ascertain whether $\widehat{\boldsymbol{\mu}}$ is admissible for sufficiently small values of p and q. Such a result would extend to the monotone incomplete setting the classical admissibility result of James and Stein [24]. Furthermore, if it is the case that these estimators are admissible for suitably small values of p and q then it will be important to explain the Stein phenomenon. Here, we have in mind extensions of the results of Brown [11] and subsequent authors. In a related direction, we are intrigued by the possibility of extending to the case of monotone incomplete samples the connection between shrinkage estimators and the heat equation as developed by Brown, *et al.* [13].

Finally, we remark that it would be useful to develop an empirical Bayes approach to motivate the estimators $\boldsymbol{m}(\widehat{\boldsymbol{\mu}}, c)$ and $\boldsymbol{m}^+(\widehat{\boldsymbol{\mu}}, c)$, hence extending the results of Efron and Morris [19].

Acknowledgments

We wish to express our deep gratitude to Steven Arnold, for we benefited greatly from extended conversations with him on the subject of the Stein phenomenon. We also are very grateful to Lawrence Brown for discussions and for his revealing lecture on shrinkage estimation and related topics in May, 2007 at Penn State University, and for the reference to his lecture notes [14].

References

[1] Anderson, T. W. (1957). Maximum likelihood estimators for a multivariate normal distribution when some observations are missing. *J. Amer. Statist. Assoc.*, **52**, 200–203.

[2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edition. Wiley, New York.

[3] Anderson, T. W. and Olkin, I. (1985). Maximum likelihood estimation of the parameters of a multivariate normal distribution. *Linear Algebra Appl.*, **70**, 147–171.

[4] Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* Wiley, New York.

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

[6] Batsidis, A. and Zografos, K. (2008). Multivariate linear regression model with elliptically contoured distributed errors and monotone missing dependent variables. *Commun. Statist. Theory Methods*, **37**, 349–372.

[7] Batsidis, A., Zografos, K., and Loukas, S. (2006). Errors in discrimination with monotone missing data from multivariate normal populations. *Comput. Statist. Data Anal.*, **50**, 2600–2634.

[8] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, second edition. Springer, New York

[9] Bhargava, R. P. (1975). Some one-sample hypothesis testing problems when there is a monotone sample from a multivariate normal population. *Ann. Inst. Statist. Math.*, **27**, 329–339.

[10] Bhatia, R. (1997). *Matrix Analysis*. Springer, New York.

[11] Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.*, **42**, 855–903.

[12] Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Lecture Notes – Monograph Series, **9**. Institute of Mathematical Statistics, Hayward, CA.

[13] Brown, L. D., DasGupta, A., Haff, L. R., and Strawderman, W. E. (2006). The heat equation and Steins identity: Connections, applications. *J. Statist. Plann. Inference*, **136**, 2254–2278.

[14] Brown, L. D. (2006). *Seminar in Statistics – Shrinkage*. University of Pennsylvania.

[15] Casella, G. and Berger, R. (1990). *Statistical Inference*. Wadsworth, Pacific Grove, CA.

[16] Chang, W.-Y. and Richards, D. St. P. (2007). Finite-sample inference with monotone incomplete multivariate normal data, I. Preprint, Penn State University.

[17] Chang, W.-Y. and Richards, D. St. P. (2008). Finite-sample inference with monotone incomplete multivariate normal data, II. Preprint, Penn State University.

[18] Eaton, M. L. and Kariya, T. (1983). Multivariate tests with incomplete data. *Ann. Statist.*, **11**, 654–665.

[19] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.*, **68**, 117–130.

[20] Fienberg, S. E. (1971). Discussion of paper by Hartley and Hocking. *Biometrics*, **27**, 813–817.

[21] Fink, A. M. and Jodeit, M. (1994). On Chebyshev's other inequality. In: *Inequalities in Statistics and Probability*, pp. 115–120, IMS Lecture Notes Monograph Series, **5**, Inst. Math. Statist., Hayward, CA.

[22] Hao, J. and Krishnamoorthy, K. (2001). Inferences on a normal covariance matrix and generalized variance with monotone missing data. *J. Multivariate Anal.*, **78**, 62–82.

[23] Hartley, H. O. and Hocking, R. R. (1971). The analysis of incomplete data. *Biometrics*, **27**, 783–823.

[24] James, W. and Stein, C. (1961). Estimation with quadratic loss. In: *Proceeding 4th Berkeley Symposium Math. Statist. Prob.*, **1**, pp. 311–319; Univ. of California Press, Berkeley, CA.

[25] Jinadasa, K. G. and D. S. Tracy (1992). Maximum likelihood estimation for multivariate normal distribution with monotone sample. *Comm. Statist. Theory Methods*, **21** (1992), 41–50.

[26] Judge, G. G. and Bock, M. E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics.* North-Holland, Amsterdam.

[27] Konno, Y. (1995). Estimation of a normal covariance matrix with incomplete data under Stein's loss. *J. Multivariate Anal.*, **52**, 308–324.

[28] Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussions). *J. Roy. Statist. Soc. Ser. B*, **34**, 1–41.

[29] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis With Missing Data*, second edition. Wiley, Hoboken, NJ.

[30] Morrison, D. F. (1971). Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data. *J. Amer. Statist. Assoc.*, **66**, 602–604.

[31] Sharma, D. and Krishnamoorthy, K. (1985). Improved minimax estimators of normal covariance and precision matrices from incomplete samples. *Calcutta Statist. Assoc. Bull.*, **34**, 23–42.

[32] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: *Proceedings 3rd Berkeley Symposium Math. Statist. Prob.*, **1**, pp. 197–206; Univ of California Press, Berkeley, CA.

[33] Stein, C. (1974). Estimation of the mean of a multivariate normal distribution. *Proc. Prague Symp. on Asymptotic Statist., 1973)*, Vol. II (J. Hájek, ed.), pp. 345–381. Charles Univ., Prague.

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

[35] Wu, L. and Perlman, M. D. (2000). Testing lattice conditional independence models based on monotone missing data. *Statist. Probab. Lett.*, **50**, 193–201.