IMPROVED ESTIMATORS FOR THE GMANOVA PROBLEM: WITH APPLICATIONS*

by

Ming Tan
Purdue University

Technical Report #89-39

Department of Statistics
Purdue University

November 1989
Revised April 1990

* Research partially supported by the National Science Foundation, Grant #DMS 8901922

◇

IMPROVED ESTIMATORS FOR THE GMANOVA PROBLEM: WITH APPLICATIONS

by

Ming Tan

Purdue University

ABSTRACT

The problem of finding classes of estimators which improve upon the usual (e.g. ML, LS) estimator of the parameter matrix in the (G)MANOVA model under (matrix) quadratic loss is considered. Unbiased estimators of risk differences for certain classes of estimators are obtained via combining integration-by-parts methods for normal and Wishart distributions, thereby extending results in Gleser [10]. A comparison of two related classes of estimators is made, and an analytic proof for risk dominance is obtained in a special case. Also considered is application to use of control variates in simulation studies to achieve better efficiency.

Key Words: GMANOVA, unbiased estimate of risk, Stein effect, shrinkage estimator, quadratic loss, matrix loss, control variates, Minimax, simulation.

◇

Improved estimators for the GMANOVA problem: with applications

Ming Tan
Department of Statistics
Purdue University

1 INTRODUCTION

The general multivariate analysis of variance (GMANOVA) model was formulated by Potthoff and Roy [22] as a generalization of multivariate analysis of variance (MANOVA) that could be applied to compare growth curves across treatment populations. This model has the form

$$U = A_1 \Theta A_2 + E, \qquad E \sim N(0, I_n \otimes \Omega), \tag{1.0}$$

where U is an $n \times r$ matrix of observed data on $r = q + p$ dependent variables, A_1 is a known $n \times m$ design matrix of rank $m < n$ and A_2 is a known $p \times r$ matrix of values of rank p obtained from p covariates (possibly including the covariate of time). The $m \times p$ parameter matrix Θ contains the unknown slope parameters which are to be estimated. Following conventions for representation of distributions of random matrices given, for example, in Muirhead [21], $E \sim N(0, I_n \otimes \Omega)$ means that the rows of E are i.i.d. r-variate normal random vectors with common mean vector 0 and covariance matrix Ω.

There has been considerable research done on the estimation (e.g. maximum likelihood) of Θ and Ω in this model (Rao [23], Geisser [8], Gleser and Olkin [12]), and particularly on developing testing procedures for various functions of Θ Kariya [16]. In addition, Hooper [15] considers invariant confidence set estimators of Θ. Kariya [16] gives an overview of the GMANOVA problem. In this paper, advantage is taken of the Stein effect to give classes of estimators improving upon the usual MLE, the maximum likelihood estimator, under various loss functions.

For any estimator $\delta = \delta(U)$ of Θ, two types of loss function will be considered: a matrix loss (Bilodeau and Kariya [5]) and a scalar loss. The matrix loss is defined to be

$$L(\delta; \Theta, \Omega) = (\delta - \Theta) Q (\delta - \Theta)' \tag{1.1}$$

where Q is a known p-dimensional positive definite matrix. For any known $t \times m$ matrix G of rank $t \leq m$, the scalar loss function

$$L^*(\delta;\Theta,\Omega) = tr[GL(\delta;\mu,\Sigma)Q'] = tr[G'G(\delta-\Theta)Q(\delta-\Theta)] \tag{1.2}$$

is the trace of $GL(\delta;\Theta,\Omega)G'$. As usual, the risk of an estimator δ is defined to be the expectation of the loss over U:

$$R(\delta;\Theta,\Omega) = E[L(\delta(U);\Theta,\Omega)], \quad R^*(\delta;\Theta,\Omega) = E[L^*(\delta(U);\Theta,\Omega)].$$

One estimator δ_1 is said to dominate another one δ_2 in risk under $L(\delta;\Theta,\Omega)$ if

$$R(\delta_2;\Theta,\Omega) - R(\delta_1;\Theta,\Omega) \geq 0 \text{ for all } \Theta \text{ and } \Omega,$$

Domination in risk under L clearly implies dominance in risk under L^*, but not necessarily conversely.

If $\hat{\Theta}$ is the MLE of Θ, it is known that $\hat{\Theta}$ is minimax under the loss

$$L^{**}(\delta;\Theta,\Omega) = \frac{L^*(\delta;\Theta,\Omega)}{R^*(\hat{\Theta};\Theta,\Omega)}.$$

It follows immediately that any estimator $\delta(U)$ which dominates $\hat{\Theta}$ in risk under the loss $L^*(\delta;\Theta,\Omega)$ is minimax for Θ under the loss $L^{**}(\delta;\Theta,\Omega)$.

Recall that A_2 has the factorization

$$A_2 = (T,0)\Gamma$$

where Γ is the $m \times m$ orthogonal and $T = (A_2A_2')^{1/2}$ in nonsingular. It is well known (Gleser and Olkin [12]) that a sufficient statistics for (Θ,Ω) is

$$(Y,X) = (A_1'A_1)^{-1}A_1'U\Gamma',$$

$$W = \Gamma U'(I_m - A_1(A_1'A_1)^{-1}A_1')U\Gamma',$$

where

$$(Y,X) \sim N((\mu,0),(A_1'A_1)^{-1}\otimes\Sigma), \quad W \sim W_{p+q}(\Sigma, n-m),$$

where (Y, X) and W are independent. Here

$$\mu = \Theta T, \quad \Sigma = \Gamma\Omega\Gamma',$$

and $W_t(\Psi, \nu)$ represents the t-dimensional Wishart distribution with degrees of freedom ν and expected value $\nu\Psi$. Also, Y and μ are $m \times p$ matrices, while X is $m \times q$. Because (Y, X, W) is sufficient for (Θ, Ω), or equivalently for (μ, Σ), it suffices to consider estimators of Θ based only on (Y, X, W). Instead, one can consider estimators $\delta(Y, X, W)$ of μ and convert these to estimators $\delta(Y, X, W)T^{-1}$ of Θ. For this purpose, the loss function (1.1) and (1.2) are replaced by

$$L(\delta; \mu, \Sigma) = (\delta - \mu)T^{-1}QT'^{-1}(\delta - \mu)',$$

$$L^*(\delta; \mu, \Sigma) = tr[G(\delta - \mu)T^{-1}QT'^{-1}(\delta - \mu)'G];$$

which simply redefines the weight matrix Q.

Consequently, the following canonical estimation problem will be considered. One observes

$$(Y, X) \sim N((\mu, 0), C \otimes \Sigma), \quad W \sim W_{p+q}(\Sigma, n - m), \tag{1.3}$$

$$(Y, X) \text{ and W statistically independent,}$$

and seeks to estimate μ under either the matrix loss function

$$L(\delta; \mu, \Sigma) = (\delta - \mu)Q(\delta - \mu)' \tag{1.4}$$

or the corresponding scalar loss

$$L^*(\delta; \mu, \Sigma) = tr[GL(\delta; \mu, \Sigma)G'] \tag{1.5}$$

where C and Q are known $p \times p$ positive definite matrices and G is a known $t \times m$ matrix of rank $t \leq m$.

The MANOVA model is the special case of (1.0) with $A_2 = I_p$ (and $q = 0$). Estimation in MANOVA using the Stein shrinkage estimators has previously been confined to improving upon the MLE of Θ (or of μ) under special assumptions either on the covariance matrix Ω or on the centering matrix Q. Scalar loss functions have previously been used. Thus, Stein [25], Efron and Morris [6] and Zheng [27] have treated the special cases $Q = \Omega = I_p$ (or $Q = \Sigma = I_p$) and $Q = I_p$, $\Omega = \sigma^2 I_p$, σ^2 an unknown positive scalar. Brown and Zidek [7] and Zidek [28] treat the case where Ω is unknown, but Q is chosen to be Ω^{-1} (a rather special situation where the problem is invariant under affine transformation). Recently, Bilodeau and Kariya [5] have obtained improvements on the MLE under a matrix loss, but with $Q = \Omega^{-1}$. The above results can (and in some cases have) be extended to GMANOVA contexts, but their applicability is limited by the strong assumptions made.

The case where Ω is unknown and Q is known has been treated primarily in the special case $m = 1$ (estimation of a normal mean vector). Berger, Bock, Brown, Casella and Gleser [3], Berger and Haff [4] and Gleser [9, 10, 11] have succeeded in developing ever wider classes of minimax estimators dominating the MLE (in scalar loss). Their results (particularly those of Gleser [10] and [11]) have recently been generalized to the MANOVA model, under both matrix and scalar loss, by Honda [14]. Some limited extensions to GMANOVA under scalar loss have also been obtained by Kubokawa and Saleh [17]. It is the purpose of the present paper to give quite general classes of dominating minimax estimators, under both scalar and matrix loss, for the GMANOVA model.

In section 2, the MANOVA model in canonical form is considered. It is shown in Section 3 that results for the GMANOVA problem can be reduced to corresponding results for the MANOVA problem by a conditioning argument (which also permits improved estimators of certain functions of the covariance matrix Ω to be obtained). Analytic comparisons among the risks of the resulting estimators are also given. In section 4, it is shown that the improved estimators developed for (the canonical form of) GMANOVA can be used to increase the efficiency of Monte Carlo simulations in multipopulation multivariate simulation experiments when control variates are used.

2. The MANOVA problem.

2.1 The problem and further canonical reduction

For the MANOVA model, $A_2 = I_p$ and $q = 0$. Consequently, the canonical model (1.3) becomes

$$Y \sim N(\mu, C \otimes \Sigma), \quad W \sim \mathcal{W}_p(\Sigma, n-m), \tag{2.1}$$

where Y and W are independent and Σ is unknown. The matrix parameter μ is to be estimated under either the matrix loss

$$L(\delta; \mu, \Sigma) = (\delta - \mu)Q(\delta - \mu)' \tag{2.2}$$

or the scalar loss

$$L^*(\delta; \mu, \Sigma) = tr[G(\delta - \mu)Q(\delta - \mu)'G']. \tag{2.3}$$

Without loss of generality this problem can be further reduced to the case where $C = I_m$, $Q = I_p$ and G is a diagonal matrix. To see this, note that the (observable) transformations

$$\tilde{Y} = F'C^{-1/2}YQ^{1/2}, \quad \tilde{W} = Q^{1/2}WQ^{1/2}$$

can be made, where $C^{1/2}$ and $Q^{1/2}$ are the unique positive definite square root of C and Q respectively, and F is an orthogonal matrix such that

$$F'C^{1/2}G'GC^{1/2}F = D_{gc} = Diag(\lambda_1(G'GC), \cdots, \lambda_m(G'GC)),$$

where as usual $\lambda_i(A)$ denotes the i^{th} characteristic root of the matrix A. Then

$$\tilde{Y} \sim N(\tilde{\mu}, I_m \otimes \tilde{\Sigma}), \quad \tilde{W} \sim \mathcal{W}_p(\tilde{\Sigma}, n-m),$$

where

$$\tilde{\mu} = F'C^{-1/2}\mu Q^{1/2}, \quad \tilde{\Sigma} = Q^{1/2}\Sigma Q^{1/2}.$$

If $\tilde{\delta}(\tilde{Y}, \tilde{W})$ is any estimator of $\tilde{\mu}$, then a corresponding estimator of μ is

$$\delta(Y, W) = C^{1/2}F\tilde{\delta}(F'C^{-1/2}YQ^{1/2}, Q^{1/2}WQ^{1/2})Q^{-1/2}$$

and

$$R(\delta;\mu,\Sigma) = C^{1/2}FE\,[(\tilde{\delta}-\tilde{\mu})(\tilde{\delta}-\tilde{\mu})']F'C^{1/2}$$
$$= C^{1/2}F\tilde{R}(\tilde{\delta};\tilde{\mu},\tilde{\Sigma})F'C^{1/2},$$
$$R^*(\delta;\mu,\Sigma) = E\,tr[GC^{\frac{1}{2}}F(\tilde{\delta}-\tilde{\mu})(\tilde{\delta}-\tilde{\mu})'F'C^{1/2}G']$$
$$= E\,tr\tilde{R}^*(\tilde{\delta};\tilde{\mu},\tilde{\Sigma}).$$

It is thus clear that $\tilde{\delta}_1$ dominates $\tilde{\delta}_2$ under the matrix risk $\tilde{R}(\tilde{\delta};\tilde{\mu},\tilde{\Sigma})$ if and only if the corresponding δ_1 dominates the corresponding δ_2 in risk under $R(\delta;\mu,\Sigma)$. Further, $\tilde{R}^*(\tilde{\delta};\tilde{\mu},\tilde{\Sigma})$ has the form of $R^*(\delta;\mu,\Sigma)$ with $\tilde{G}=D_{gc}^{1/2}$ replacing G and Q replaced by I_p.

Consequently, in the remainder of this section it is assumed that $C = I_m$, $Q = I_p$, $G = Diag(g_1,\cdots,g_m)$,

$$Y \sim N(\mu, I_m \otimes \Sigma), \quad W \sim W_p(\Sigma, n-m), \quad Y \text{ and } W \text{ independent} \tag{2.4}$$

$$L(\delta;\mu,\Sigma) = (\delta-\mu)(\delta-\mu)', \tag{2.5}$$

$$L^*(\delta;\mu,\Sigma) = tr[GL(\delta;\mu,\Sigma)G']. \tag{2.6}$$

Note that (2.4) implies that the rows Y_i' of Y are independent, $Y_i \sim N(\mu,\Sigma)$, where μ_i is the i^{th} row of μ.

2.2 Unbiased estimator of the difference in risks.

Suppose that δ is an arbitrary estimator of μ (such an estimator can be written as $\delta(Y,W) = Y - t(Y,W)$). Let

$$\Delta = \Delta(\mu,\Sigma) = R(\delta;\mu,\Sigma) - R(\delta_0;\mu,\Sigma),$$

where $R(\delta;\mu,\Sigma) = E(\delta-\mu)(\delta-\mu)'$, and $\delta_0 = Y$.

The estimators that we will consider are of the form

$$\delta(Y,W) = Y - h(Y,W) - \frac{2}{n^*-p-1}r(Y,W),$$

where $n^* = n-m$. Using the technique in Gleser [10], an unbiased estimator of Δ can be obtained for estimators δ having the form $Y - t(Y,W)$, where

$$t(Y,W) = h(Y,W) + \frac{2}{n^*-p-1}r(Y,W), \tag{2.7}$$

$h(Y,W) = (h_1(Y,W), \ldots, h_m(Y,W))'$ is a given matrix-valued function of Y and W, and $r(Y,W)$ is defined through $h(Y,W)$ by

$$r(Y,W) = \begin{pmatrix} r_1'(Y,W) \\ \vdots \\ r_m'(Y,W) \end{pmatrix}, \quad r_j(Y,W) = \begin{pmatrix} r_j^{(1)}(Y,W) \\ \vdots \\ r_j^{(m)}(Y,W) \end{pmatrix} \tag{2.8}$$

and

$$r_j^{(i)}(Y,W) = \frac{\partial (Wh_j(Y,W))_i}{\partial w_{ii}} + \frac{1}{2} \sum_{\ell \neq i} \frac{\partial (Wh_j(Y,W))_\ell}{\partial w_{i\ell}}, \tag{2.9}$$

Here for any vector u (column or row), $(u)_i$ always denotes the ith element of u. Also define as usual the matrix

$$J_{\gamma(u)}(u) = \left(\frac{\partial \gamma_j(u)}{\partial u_i} \right) = \frac{\partial \gamma(u)}{\partial u},$$

where u, γ are k-dimensional vectors and γ is a function of u, $\gamma(u) = (\gamma_1(u), \ldots, \gamma_k(u))'$. And define the matrices

$$J = J(Y,W) = (J_{ij}) = (J_{h_j(Y,W)}(Y_i)),$$

$$W \circ J = (WJ_{ij}) = \begin{pmatrix} WJ_{11} & \cdots & WJ_{1m} \\ \vdots & & \vdots \\ WJ_{m1} & \cdots & WJ_{mm} \end{pmatrix}$$

and for any matrix $N = (N_{ij})_{mp \times mp}$ where N_{ij} are submatrices, $i, j = 1, \ldots, m$, let

$$T_R N = (trN_{ij})_{m \times m}.$$

Then the following result generalizes Theorem 1 in Gleser [10]

Theorem 2.2.1. Let $h_j(Y,W)$ $(i = 1, \ldots, m)$ satisfy the regularity conditions of Lemma A.2 (See Appendix) and let $t(Y,W)$ be defined by (2.7)–(2.9). Define the estimator

$$\delta(Y,W) = Y - t(Y,W).$$

Then if $\delta(Y,W)$ has finite risk (each element of the risk matrix is finite),

$$\Delta = R(\delta, \mu, \Sigma) - R(\delta_0, \mu, \Sigma) = EM(Y,W),$$

where

$$M(Y,W) = t(Y,W)t'(Y,W) - \frac{1}{n^*-p-1} T_R(W \circ J(Y,W)) - \frac{1}{n^*-p-1} T_R(W \circ J(Y,W))'$$

PROOF: First note that $E\ tr\ t(Y,W)t(Y,W)' < \infty$ implies $EL(\delta(Y,W);\mu,\Sigma) < \infty$. Now

$$\Delta = E\ [t(Y,W)t'(Y,W)] - E\ [(Y-\mu)t'(Y,W)] - E\ [t(Y,W)(Y-\mu)']. \tag{2.10}$$

The (i,j) element of $E\ [(Y-\mu)t'(Y,W)]$ is

$$E\ [(Y_i-\mu_i)'t_j(Y,W)] = E\ [(Y_i-\mu_i)'h_j(Y,W)] + \frac{2}{n^*-p-1} E[(Y_i-\mu_i)'r_j(Y,W)],$$

By Haff's identity (Appendix, Lemma A.2)

$$\begin{aligned} E\ [(Y_i-\mu_i)'h_j] &= E\ [trW^{-1}(Y_i-\mu_i)(Wh_j(Y,W))'] \\ &= \frac{1}{n^*-p-1} E\ [tr\Sigma^{-1}T_j(Y_i,W)] - \frac{2}{n^*-p-1} E\ [D_{(1/2)}T_j(Y_i,W)] \end{aligned}$$

where $T_j(Y_i,W) = (Y_i-\mu_i)(Wh_j(Y,W))'$, the derivative

$$\begin{aligned} D_{(1/2)}T_j(Y_i,W) &= D_{(1/2)}(Y_i-\mu_i)(Wh_j(Y,W))' \\ &= r_j'(Y,W)(Y_i-\mu_i) \\ &= (Y_i-\mu_i)'r_j(Y,W), \end{aligned}$$

and the i^{th} element of r_j is defined as

$$\begin{aligned} r_j^{(i)}(Y,W) &= \frac{\partial(Wh_j(Y,W))_i}{\partial w_{ii}} + \frac{1}{2}\sum_{\ell \neq i} \frac{\partial(Wh_j(Y,W))_\ell}{\partial w_{i\ell}} \\ &= \frac{1}{2}\sum_{\ell=1}^{p}(1+\delta_{i\ell})\frac{\partial(Wh_j(Y,W))_\ell}{\partial w_{i\ell}} \end{aligned}$$

Here, δ_{ij} is the Kronecker delta ($\delta_{ij}=1$ if $i=j$; 0, otherwise). Denote $c = 2/(n^*-p-1)$ throughout this section,

$$\begin{aligned} E\ [(Y_i-\mu_i)'t_j(Y,W)] &= \frac{1}{n^*-p-1} E\ tr[\Sigma^{-1}T_j(Y_i,W)] - c\,E\ [(Y_i-\mu_i)'r_j(Y,W)] \\ &\quad + c\,E\ [(Y_i-\mu_i)'r_j(Y,W)] \\ &= \frac{1}{n^*-p-1} E\ tr[\Sigma^{-1}T_j(Y_i,W)]. \end{aligned}$$

Integration by parts (see Appendix, Lemma A.1) gives

$$\begin{aligned} E\ [(Y_i-\mu_i)'t_j(Y,W)] &= \frac{1}{n^*-p-1} E\ tr[\Sigma^{-1}T_j(Y_i,W)] \\ &= \frac{1}{n^*-p-1} E\ tr[\Sigma^{-1}(Y_i-\mu_i)(Wh_j(Y,W))'] \\ &= \frac{1}{n^*-p-1} E\ tr[WJ_{h_j}(Y_i)]. \end{aligned}$$

Consequently,

$$E\ [(Y-\mu)t'(Y,W)] = \frac{1}{n^*-p-1} E\ [T_R W \circ J(Y,W)]$$

Combining this result with (2.10) yields the needed result.

2.3 Classes of improved estimators

2.3.1 Modified estimators using information between coordinates and repetitions.

The first class of estimators considered are based on a matrix-valued version of a function considered in Gleser [10] for the case $m = 1$. That is,

$$h(Y,W) = \frac{b(W)D_c}{trYW^{-1}Y'} YW^{-1} \tag{2.11}$$

where $D_c = Diag(c_1, c_2, \cdots, c_m)$, $c_i > 0$, for $i = 1, \cdots, m$ and $b(W)$ is a positive scalar function of W which is continuously differentiable as a function of the $p(p+1)/2$ free elements of W. Then resulting class of improved estimators of μ is

$$\delta(Y,W) = Y - \frac{b(W)D_c}{trYW^{-1}Y'}(MYW^{-1} + cYU'(W)),$$

where $M = I + cYW^{-1}Y'/trYW^{-1}Y'$, $U(W)$ is defined as

$$U(W) = \left(\frac{1+\delta_{ij}}{2} \frac{\partial}{\partial w_{ij}} \log b(W) \right). \tag{2.12}$$

To see this, note for this choice of $h(Y,W)$,

$$(Wh_k(Y,W))_j = \frac{c_k b(W)}{trYW^{-1}Y'} y_{kj},$$

where $(y_{ij}) = Y$. Also note that

$$\frac{\partial}{\partial w_{ij}} Y'_\alpha W^{-1} Y_\alpha = -(2-\delta_{ij})(Y'_\alpha W^{-1})_i (Y'_\alpha W^{-1})_j$$

and

$$\frac{\partial (Wh_k(Y,W))_j}{\partial w_{ij}} = \frac{c_k \partial b(W)}{\partial w_{ij}} \frac{1}{trYW^{-1}Y'} y_{kj} + \frac{(2-\delta_{ij})c_k b(W)}{(trYW^{-1}Y')^2} \sum_{\alpha=1}^{m} (Y'_\alpha W^{-1})_i (Y'_\alpha W^{-1})_j y_{kj}.$$

Thus

$$\begin{aligned} r_k^{(i)}(Y,W) &= \frac{c_k b(W)}{trYW^{-1}Y'} \Big\{ \frac{1}{2} \sum_{j=1}^{p} (1+\delta_{ij}) \frac{\partial \log b(W)}{\partial w_{ij}} y_{kj} \\ &+ \frac{1}{2} \sum_{j=1}^{p} \frac{(1+\delta_{ij})(2-\delta_{ij})}{trYW^{-1}Y'} \sum_{\alpha=1}^{m} (Y'_\alpha W^{-1})_i (Y'_\alpha W^{-1})_j y_{kj} \Big\} \\ &= \frac{c_k b(W)}{trYW^{-1}Y'} \left[\sum_{j=1}^{p} U_{ij} y_{kj} + \frac{1}{trYW^{-1}Y'} \sum_{\alpha=1}^{m} (Y'_\alpha W^{-1})_i Y'_\alpha W^{-1} Y_k \right], \end{aligned}$$

and

$$r_k(Y,W) = \frac{c_k b(W)}{trYW^{-1}Y'} \left[\begin{pmatrix} (YU)'_{k1} \\ \vdots \\ (YU)'_{km} \end{pmatrix} + \frac{1}{trYW^{-1}Y'} \sum_{\alpha=1}^{m} W^{-1} Y_\alpha Y'_\alpha W^{-1} Y_k \right],$$

$$r(Y,W) = \begin{pmatrix} r'_1 \\ \vdots \\ r'_m \end{pmatrix} = \frac{b(W)D_c}{trYW^{-1}Y'} (YU'(W) + \frac{YW^{-1}Y'YW^{-1}}{trYW^{-1}Y'}).$$

It follows that

$$\begin{aligned} t(Y,W) &= h(Y,W) + \frac{2}{n^*-p-1} r(Y,W) \\ &= \frac{b(W)D_c}{trYW^{-1}Y'} [(I + \frac{2}{n^*-p-1} \frac{YW^{-1}Y'}{trYW^{-1}Y'})YW^{-1} + \frac{2}{n^*-p-1} YU'(W)] \\ &= \frac{b(W)D_c}{trYW^{-1}Y'} (MYW^{-1} + cYU'(W)), \end{aligned}$$

Finally,

$$\delta = Y - t(Y,W) = Y - \frac{b(W)D_c}{trYW^{-1}Y'} (MYW^{-1} + cYU'(W)), \tag{2.13}$$

where $U(W)$ is defined by (2.12).

Theorem 2.3.1. Suppose that $h(Y,W)$ satisfies the regularity conditions for establishing the Wishart identity (Appendix, Lemma A.2). Also, assume that $U'(W)U(W) \leq W^{-2}$, where $U(W)$ is defined by (2.12) and that

$$0 \leq b(W) \leq \frac{2(p-2a)}{c_{max}} \frac{n^*-p-1}{(n^*-p+3)^2} \lambda_{\min}(W), \tag{2.14}$$

where $p > 2a$ and $a = (c_{max}/c_{min})^{1/2}$, $c_{max} = \max_{1 \leq i \leq m} (c_i)$, $c_{min} = \min_{1 \leq i \leq m} (c_i)$ and $\lambda_{\min}(W)$ is the smallest eigenvalue of W. Then the estimator defined by (2.10) dominates $\delta_0(Y, W) = Y$ in risk under the matrix loss (2.5).

NOTE: (i) The following common choices of $b(W)$ satisfy the condition $U'U \leq W^{-2}$:

$$b(W) = c(trW^{-1})^{-1} \quad \text{and} \quad b(W) = c\lambda_{\min}(W),$$

where c is a constant (Gleser [10]).

(ii) If we take $D_c = I_m$, then (2.14) becomes the more familiar condition

$$0 \leq b(W) \leq 2(p-2)\frac{n^*-p-1}{(n^*-p+3)^2}\lambda_{\min}(W).$$

PROOF: By Theorem 2.1.1, the unbiased estimator of the difference in risks is

$$M(Y, W) = t(Y, W)t'(Y, W) - \frac{c}{2}T_R(W \circ J(Y, W)) - \frac{c}{2}[T_R(W \circ J(Y, W))]' \tag{2.15}$$

Then for any m-dimensional vector $a \neq 0$, by Lemma A.3,

$$a't(Y, W)t'(Y, W)a \leq \frac{b^2(W)}{(trYW^{-1}Y')^2}(\sqrt{a'D_cMYW^{-2}Y'MD_ca} + c\sqrt{a'D_cYU'UY'D_ca})^2. \tag{2.16}$$

Denote $A_0 = YW^{-1}Y'$, then $M = I + cA_0/trA_0$. Note that both A_0 and M are symmetric matrices. Then $M^2 = MM' = I + 2cA_0/trA_0 + c^2A_0A_0/(trA_0)^2$. Also note the following easy proven facts:

(i) For any p.s.d. matrix S,

$$S \leq (trS)I \quad \text{and} \quad (I + c\frac{S}{trS})S(I + c\frac{S}{trS}) \leq (1+c)^2 S$$

(ii) $W^{-2} \leq \lambda_{min}^{-1}(W)W^{-1}$

Therefore

$$\begin{aligned} MYW^{-2}Y'M &\leq \lambda_{min}^{-1}(W)MYW^{-1}Y'M \\ &= \lambda_{min}^{-1}(W)\left[(I + c\frac{A_0}{trA_0})A_0(I + c\frac{A_0}{trA_0})\right] \\ &\leq (1+c)^2\lambda_{min}^{-1}(W)A_0. \end{aligned}$$

and

$$(a'D_cMYW^{-2}Y'MD_ca) \leq (1+c)^2\lambda_{min}^{-1}(W)a'D_cYW^{-1}Y'D_ca.$$

Combining this with (2.6) and the assumption that $UU' \leq W^{-2}$ yields

$$\begin{aligned} a't(Y,W)t'(Y,W)a &\leq \frac{b^2(W)}{(trYW^{-1}Y')^2}(1+c+c)^2\lambda_{min}^{-1}(W)a'D_cYW^{-1}Y'D_ca \\ &= \frac{b^2(W)}{trYW^{-1}Y'}(1+2c)^2\lambda_{\min}^{-1}(W)a'D_c\frac{YW^{-1}Y'}{trYW^{-1}Y'}D_ca. \end{aligned}$$

Therefore,

$$\begin{aligned} t(Y,W)t'(Y,W) &\leq \frac{b^2(W)}{trYW^{-1}Y'}(1+2c)^2\lambda_{\min}^{-1}(W)D_c\frac{YW^{-1}Y'}{trYW^{-1}Y'}D_c \\ &\leq \frac{b^2(W)}{trYW^{-1}Y'}(1+2c)^2\lambda_{\min}^{-1}(W)D_c^2 \end{aligned} \tag{2.17}$$

On the other hand, the (i,j) element of $T_RW \circ J(Y,W)$ is

$$tr[WJ_{h_j}(Y_i)] = tr[J_{(Wh_j)}(Y_i)].$$

Because

$$\begin{aligned} Wh_j &= \frac{c_jb(W)}{trYW^{-1}Y'}Y_j \\ tr[J_{Wh_j}(Y_i)] &= \begin{cases} \frac{b(W)c_i}{trYW^{-1}Y'}p - 2b(W)\frac{Y_i'W^{-1}Y_ic_i}{(trYW^{-1}Y')^2} & \text{if } i = j \\ -2\frac{b(W)}{(trYW^{-1}Y')^2}Y_i'W^{-1}Y_jc_j & \text{if } i \neq j, \end{cases} \end{aligned} \tag{2.18}$$

it follows that

$$\begin{aligned} T_R(W \circ J(Y,W)) &= \frac{b(W)}{trYW^{-1}Y'}\left(pI_m - 2\frac{YW^{-1}Y'}{trYW^{-1}Y'}\right)D_c \\ &= \frac{b(W)}{trYW^{-1}Y'}(pI_m - 2\frac{A_0}{trA_0})D_c. \end{aligned}$$

By Lemma A.4 (see Appendix),

$$\frac{A_0}{trA_0}D_c + D_c\frac{A_0}{trA_0} \leq 2D_c a$$

and combining this with (2.17) and (2.18) yields

$$\begin{aligned}
M(Y,W) \leq & \frac{b(W)}{trYW^{-1}Y'}((1+2c)^2\lambda_{\min}^{-1}(W)b(W)D_c^2 - \frac{1}{n^*-p-1}(pI_m - 2\frac{A_0}{trA_0})D_c \\
& - \frac{1}{n^*-p-1}D_c(pI_m - 2\frac{A_0}{trA_0})) \\
= & \frac{b(W)}{trYW^{-1}Y'}((1+2c)^2\lambda_{\min}^{-1}(W)b(W)D_c^2 - \frac{2}{n-p-1}pD_c \\
& + \frac{2}{n-p-1}(\frac{A_0}{trA_0}D_c + D_c\frac{A_0}{trA_0})) \\
\leq & \frac{cb(W)D_c^{\frac{1}{2}}}{trYW^{-1}Y'}\left[\frac{(n^*-p+3)^2}{n^*-p-1}b(W)\lambda_{\min}^{-1}(W)D_c - 2(p-2a)I_m\right]D_c^{\frac{1}{2}}. \qquad (2.19) \\
\leq & \; 0.
\end{aligned}$$

Q.E.D.

Observe that

$$\begin{aligned}
tr\left[G^2D_c(pI - 2\frac{YW^{-1}Y'}{trYW^{-1}Y'})\right] &= \sum_{i=1}^{m} c_i g_i^2\left(p - 2\frac{Y_i'W^{-1}Y_i}{trYW^{-1}Y'}\right) \\
&\geq c_{\min}g_{min}^2\sum_{i=1}^{m}\left(p - 2\frac{Y_i'W^{-1}Y_i}{trYW^{-1}Y'}\right) \\
&= c_{\min}g_{min}^2(pm-2).
\end{aligned}$$

where $g_{min} = \min_{1\leq i\leq m}(g_i)$, hence, from the first inequality of (2.16),

$$\begin{aligned}
tr[G^2M(Y,W)] \leq & \frac{b(W)}{(n^*-p-1)trYW^{-1}Y'}[trD_c^2G^2\frac{(n^*-p+3)^2}{n^*-p-1}b(W)\lambda_{\min}^{-1}(W) \\
& - 2c_{min}g_{min}^2(pm-2)] \\
\leq & \; 0
\end{aligned}$$

provided that

$$b(W) \leq \frac{2c_{min}g_{min}^2(pm-2)}{(c_1^2g_1^2 + \cdots + c_m^2g_m^2)}\left[\frac{n^*-p-1}{(n^*-p+3)^2}\right]\lambda_{\min}(W).$$

It follows that the condition on $b(W)$ in Theorem 2.3.1 can be weakened in order to obtain risk dominance under the scalar loss function L^*. We have

Corollary 2.3.1. Let $pm \geq 3$, and let the conditions on $b(W)$ in Theorem 2.3.1 be retained. Then $\delta = Y - t(Y,W)$ defined by (2.13) dominates $\delta_0 = Y$ in risk with respect to $L^*(\delta;\mu,\Sigma)$ provided

$$b(W) \leq \frac{2c_{min}g_{min}^2(pm-2)}{(c_1^2g_1^2 + \cdots + c_m^2g_m^2)}\left[\frac{n^*-p-1}{(n^*-p+3)^2}\right]\lambda_{\min}(W).$$

Another class of improved estimates can be obtained by considering a matrix version of a function $h(Y,W)$ proposed by Gleser [11] (see also Berger and Haff [4]). Let

$$h(Y,W) = c_0\alpha(W)s(tr[YW^{-1}Y'])YW^{-1}, \tag{2.20}$$

where $c_0 \geq 0$, the scalar function $s(\cdot)$ maps $[0,\infty)$ into $(0,\infty)$ and is continuous and differentiable, and the scalar function $\alpha(W)$ is nonnegative, continuous, and differentiable (with respect to the elements of W).

Then, using (2.8), (2.9), the usual calculus gives

$$r(Y,W) = c\alpha(W)[s(v)YU(W) - s'(v)YW^{-1}Y'YW^{-1}],$$

where

$$v = tr[YW^{-1}Y'],$$

$$U(W) = (U_{ij}(W)),\ U_{ij}(W) = \begin{cases} \frac{\partial \log\alpha(W)}{\partial W_{ii}} & i = j \\ \frac{1}{2}\frac{\partial \log\alpha(W)}{\partial W_{ij}} & i \neq j. \end{cases}$$

Also,

$$\begin{aligned} T_R[W \circ J(Y,W)] &= (tr[WJ_{h_j}(Y_i)]) \\ &= c\alpha(W)(ps(v)I_m + 2s'(v)YW^{-1}Y'), \end{aligned}$$

In consequence,

$$\begin{aligned} t(Y,W) &= h(Y,W) + \frac{2}{n^*-p-1}r(Y,W) \\ &= c_0\alpha(W)s(v)[M_gYW^{-1} + \frac{2}{n^*-p-1}YU(W)] \end{aligned} \tag{2.21}$$

where

$$M_g = I - \frac{2}{n^* - p - 1}\frac{s'(v)}{s(v)}YW^{-1}Y'$$
$$= I + c_1\frac{A_0}{trA_0}$$

with

$$c_1 = -\frac{2}{n^* - p - 1}\frac{s'(v)}{s(v)}v.$$

The new class of improved estimators is given by

$$\delta(Y, W) = Y - t(Y, W) \tag{2.22}$$

where $t(Y, W)$ is defined in (2.21).

Theorem 2.3.2. Suppose that $\delta(Y, W) = Y - t(Y, W)$ has finite risk, and that $s(v)$ is nonincreasing, $U(W)U'(W) \leq W^{-2}$, and

$$\alpha(W) \leq \frac{\lambda_{\min}(W)}{n^* - p - 1}.$$

Define $g(v) = cv\left[(n-p+1)s(v) - 2s'(v)v)^2 - 2(ps(v) + 2vs'(v)\right]/(n-p-1)$. Then if $g(v) \leq 0$ for all $v \geq 0$, $\delta(Y, W)$ defined by (2.22) dominates Y in risk under the matrix loss L.

PROOF: Using the same idea used in proving Theorem 2.3.1, and noticing that now

$$M = M_g = I + c_1\frac{A_0}{trA_0}, \qquad c_1 = -\frac{2}{n^* - p - 1}\frac{s'(v)}{s(v)}v \geq 0 \qquad (\text{since } s'(v) \leq 0),$$

then

$$t(Y, W)t'(Y, W) \leq c_0^2\alpha^2(W)vs^2(v)(1 + c_1 + \frac{2}{n^* - p - 1})^2\lambda_{\min}^{-1}(W)I_m.$$

Since $\alpha(W) \leq \lambda_{\min}(W)/(n-p-1)$, it follows that

$$t(Y, W)t'(Y, W) \leq c_0\alpha(W)s(v)\left(\frac{c_0 s(v)v}{n^* - p - 1}(1 + c_1 + \frac{2}{n^* - p - 1})^2\right)$$

Then

$$
\begin{aligned}
M(Y,W) &= t(Y,W)t'(Y,W) - \frac{1}{n^* - p - 1} E[(T_R(W \circ J(Y,W)) + T_R(W \circ J(Y,W))']\\
&\le \frac{c_0 \alpha(W) s(v)}{n^* - p - 1}[cs(v)v(1 + c_1 + \frac{2}{n^* - p - 1})^2 - 2(p + 2v\frac{s'(v)}{s(v)})]\\
&= \frac{c_0 \alpha(W)}{n^* - p - 1}[\frac{c_0 v}{(n^* - p - 1)^2}((n^* - p + 1)s(v) - 2s'(v)v)^2 - 2(ps(v) + 2vs'(v))]\\
&= \frac{c_0 \alpha(W)}{n^* - p - 1} g(v)\\
&\le 0.
\end{aligned}
$$

Q.E.D.

Furthermore, the proof of Corollary 1 in Gleser [11] yields

Corollary 2.3.2. Under the assumptions of Theorem 2.3.1 suppose that $s(v)$ satisfies "condition h" of Berger and Haff [4]: that is, $s(v)$ is continuous and piecewise differentiable. Also suppose that for all $v \ge 0$,

$$0 \le vs(v) \le 1, \quad s(v) + vs'(v) \ge 0, \quad s'(v) \le 0.$$

Then $\delta(Y, W)$ dominates Y in risk under the matrix loss L if $\delta(Y, W)$ has finite risk and c_0 satisfies

$$0 < c_0 \le \frac{2(p-2)(n^* - p - 1)}{(n^* - p + 3)^2}.$$

Remark: If the scalar loss is considered, similar result to Corollary 2.3.1 is immediate with $c_i = 1$.

2.3.2 Stein estimators applied row-wise

The estimators in Theorems 2.3.1 and 2.3.2 can be regarded as treating Y as if it were a pm-dimensional vector. Except for the use of different constants c_i to weight each row of Y in Theorem 2.3.1, each element of Y is adjusted by the same function of Y and W to produce a better estimator of μ. Instead, since the rows Y_i' of Y are independent, Gleser's [10, 11] approach can be applied row-wise to Y. That is,

$$h(Y, W) = (h_1(Y, W), \cdots, h_m(Y, W))',$$

where

$$h_k(Y,W) = h_k(Y_k,W) = \frac{c_k b(W)}{Y_k' W^{-1} Y_k} W^{-1} Y_k, \quad k = 1, \cdots, m. \tag{2.23}$$

If

$$0 < b(W) c_k \leq 2(p-2) \left[\frac{n^* - p - 1}{(n^* - p + 3)^2} \right] \lambda_{min}(W), \quad k = 1, \cdots, m,$$

and

$$U'(W)U(W) \leq W^{-2},$$

where $U(W)$ is defined by (2.12), then the estimator of μ obtained row-wise dominates Y in risk under the usual square error loss. It is then easy to see that such an estimator dominates Y in risk under the scalar loss $L^*(\delta;\mu,\Sigma)$. Although row-wise domination is obviously necessary for domination under the matrix loss $L(\delta;\mu,\Sigma)$, it is not clearly sufficient for such domination. Consequently the following theorem is of some interest. Before stating this theorem, note that the estimator obtained from (2.23) is

$$\delta(Y,W) = Y - t(Y,W) = Y - (t_1(Y_1,W), \cdots, t_k(Y_m,W))', \tag{2.24}$$

where

$$t_k(Y_k,W) = \frac{c_k b(W)}{Y'kW^{-1}Y_k} [(1+c)W^{-1}Y_k + cU(W)Y_k]$$

and $c = 2/(n^* - p - 1)$.

Theorem 2.3.3 If $h_i(Y_i,W)$ satisfies the regularity conditions for establishing the Wishart identity (Appendix, Lemma A.2), and if $U'(W)U(W) \leq W^{-2}$, where $U(W)$ is defined by (2.12) and also

$$0 \leq b(W) \leq \frac{2(p-2)}{mc_{max}} \frac{n^* - p - 1}{(n^* - p + 3)^2} \lambda_{\min}(W), \tag{2.25}$$

Then the estimator defined by (2.24) dominates $\delta_0(Y,W) = Y$ in risk under the matrix loss $L(\delta;\mu,\Sigma)$.

PROOF: Using the the same argument in proving Theorem 2.3.2, by Lemma A.3 and the assumption that $UU' \leq W^{-2}$, we have

$$\begin{aligned} a't(Y,W)t'(Y,W)a &\leq b^2(W)(1+c+c)^2 a' D_c D_v^{-1} Y W^{-2} Y' D_v^{-1} D_c a \\ &\leq b^2(W)(1+2c)^2 \lambda_{\min}^{-1}(W) a' D_c D_v^{-1} Y W^{-1} Y' D_v^{-1} D_c a, \end{aligned}$$

where

$$D_v = Diag(v_1, \cdots, v_m), \quad v_k = Y_k' W^{-1} Y_k.$$

Staightforward matrix algebra however gives

$$tr\left[D_v^{-1/2} Y W^{-1} Y' D_v^{-1/2}\right] = m$$

Thus $\Delta^{-1/2} Y W^{-1} Y' \Delta^{-1/2} \leq m I_m$, and

$$t(Y,W)t'(Y,W) \leq m b^2(W) \left(\frac{n-p+3}{n-p-1}\right)^2 \lambda_{\min}^{-1}(W) D_c D_v^{-1} D_c \tag{2.26}$$

On the other hand, the (i,j) element of $T_R[W \circ J(Y,W)]$ is

$$tr[W J_{h_j}(Y_i)] = tr[J_{(Wh_j)}(Y_i)]$$

Calculus then gives

$$tr J_{(Wh_j)}(Y_i) = \begin{cases} (p-2)\frac{b(W)c_i}{Y_i' W^{-1} Y_i} & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.27}$$

Thus

$$T_R[W \circ J(Y,W)] = b(W)(p-2) D_c D_v^{-1}$$

So by (2.26) and (2.27)

$$\begin{aligned} M(Y,W) &\leq \frac{b(W) D_c D_v^{-1}}{n^* - p - 1} \left[m \frac{(n^* - p + 3)^2}{n^* - p - 1} b(W) \lambda_{\min}^{-1}(W) D_c - 2(p-2) I_m \right] \quad (2.28) \\ &\leq 0. \end{aligned}$$

Q.E.D.

Remark 1. The common choices of $b(W)$ still work, e.g., $b(W)$ can be taken as one of the choices

$$b(W) = (n^* - p - 1)^{-1} \lambda_{\min}(W), \quad b(W) = (n^* - p - 1)^{-1} [tr W^{-1}]^{-1}$$

Remark 2. Although we have obtained improved estimators, the analytic evaluation of the potential gain (or savings) seems untractable, and numerical calculation or simulation of risks becomes a necessary resort for this purpose.

3 GMANOVA

3.1 Improved Classes of Estimates.

3.1.1 Distributional results and a heuristic estimator.

We now turn to the GMANOVA problem defined in (1.3)–(1.5). Using Theorem 3.2.10 in Muirhead [21] (and adopting the notations used there as well), it is easy to see that the following holds.

Lemma 3.1.1. If $(Y, X) \sim N((\mu, 0), \Sigma)$,

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \sim \mathcal{W}_{p+q}(n^*, \Sigma),$$

and (Y, X) is independent of W, then

$$Y|X \sim N(\mu + X\beta', I_m \otimes \Sigma_{11.2}), \quad X \sim N(0, I_m \otimes \Sigma_{22}),$$

$$W_{11.2} \sim \mathcal{W}_p(k, \Sigma_{11.2}), \quad k = n^* - q, \quad B' = W_{22}^{-1} W_{21} | W_{22} \sim N(\beta', W_{22}^{-1} \otimes \Sigma_{11.2}),$$

where

$$\beta = \Sigma_{12}\Sigma_{22}^{-1}, \quad \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \quad W_{11.2} = W_{11} - W_{12}W_{22}^{-1}W_{21},$$

and $W_{11.2}$ is independent of (Y, X), W_{12}, W_{22}, and, consequently, B.

As is well known $Z = Y - XB'$ is the ML and unbiased estimator of μ. From this Lemma, it is easy to see that

$$Z = Y - XB' = X - X\beta' - X(B - \beta)',$$

and $Y - X\beta'$ and $X(B - \beta)'$ are independent. Using Lemma 3.1.1,

$$\text{Cov}\,(Z) = (I_m + XW_{22}^{-1}X') \otimes \Sigma_{11.2},$$

we then have

$$Z|X, W_{22} \sim N(\mu, (I_m + XW_{22}^{-1}X') \otimes \Sigma).$$

Let $A = I_m + XW_{22}^{-1}X'$, $Z^* = A^{-1/2}Z$, $\mu^* = A^{-1/2}\mu$, then

$$Z^*|X, W_{22} \sim N(\mu^*, I_m \otimes \Sigma). \tag{3.2}$$

However, from Lemma 3.1.1

$$Y - X\beta' \sim N(\mu, I_m \otimes \Sigma_{11.2}).$$

If β is known, then a direct application of Theorem 2.2.1 gives a class of improved estimators. Since β is typically unknown, we substitute its unbiased estimator B for β in those improved estimators and hence obtain our first class of estimators of μ:

$$\delta_1(Z, W) = Z - t(Z, W), \tag{3.3}$$

where $W = W_{22.1}$ (also $\Sigma = \Sigma_{22.1}$) by suppressing the subscript for notational brevity, and

$$t(Z, W) = \frac{b(W)}{trZW^{-1}Z'} \ (MZW^{-1} + \frac{2}{k-p-1} ZU(W)) \tag{3.4}$$

with $k = n^* - q$ and

$$M = I + \frac{2}{k-p-1} \ \frac{ZW^{-1}Z'}{trZW^{-1}Z'}.$$

A natural question is if these estimators continue to dominate $\delta_0 = Z$ in risk. In other words since the distribution of Z is after all elliptical, i.e., of the form $f((Z-\mu)\Sigma_{11.2}^{-1}(Z-\mu)')$, and the problem is whether the estimator improving on the normal mean still better than the MLE for this kind of distributional departure from normality. The answer is positive at least under the total squared error loss and in fact we have

Theorem 3.1.1. Suppose the regularity conditions in Theorem 2.3.1 hold, and if $pm \geq 2$, $p \geq 2\, g_{max}/g_{min}$ and

$$b(V) \leq \frac{2g_{min}^2(pm-2)}{tr[G^2]} \frac{n^*-p-q-1}{(n^*-p-q+3)^2} \lambda_{\min}(V), \quad \text{for any matrix } V \geq 0. \tag{3.6}$$

Then the estimators defined by (3.3)-(3.5) dominate Z in risk (under L^*, the total squared error loss.)

NOTE. If $G = I$ is taken, the condition is then simply $pm \geq 2$ and $p \geq 2$ and (3.6) becomes

$$b(V) \leq \frac{2(pm-2)}{m} \frac{n^* - p - q - 1}{(n^* - p - q + 3)^2} \lambda_{\min}(V), \qquad \text{for any matrix } V \geq 0.$$

PROOF: By following the argument in Theorem 2.1.1., we can find an unbiased estimator of the difference in risks. Let

$$h(Z, W) = \frac{b(W)}{trZW^{-1}Z'} ZW^{-1}, \quad t(Z, W) = h(Z, W) + \frac{2}{k-p-1} r(Z, W),$$

then

$$\begin{aligned} \Delta_1 &= R(\delta_1; \mu, \Sigma) - R(Z; \mu, \Sigma) \\ &= E(t(Z, W)t'(Z, W) - (Z - \mu)t'(Z, W) - t(Z, W)(Z - \mu)'). \end{aligned} \tag{3.5}$$

The (i, j) element of $E[(Z - \mu)t'(Z, W)]$ is

$$E[(Z_i - \mu_i)'h_j(Z, W)] = \frac{1}{k-p-1}\ E\ tr[\Sigma^{-1}T_j(Z_i, W)] - \frac{2}{k-p-1}\ E\ tr[D_{(\frac{1}{2})}T_j(Z_i, W)],$$

where $T_j(Z_i, W) = (Z_i - \mu_i)(Wh_j(Z, W))'$. So if we let r be defined as in (2.4) and (2.5) , then

$$\begin{aligned} E[(Z_i - \mu_i)'t_j(Z, W)] &= \frac{1}{k-p-1}\ E\ tr[\Sigma^{-1}T_j(Z_i, W)] \\ &= \frac{1}{k-p-1}\ E\ [(Z_i - \mu_i)'\Sigma^{-1}Wh_j(Z, W)] \end{aligned} \tag{3.7}$$

Finally integration by parts (see Appendix A.5 for details and verifications) gives

$$E[(Z - \mu)t'(Z, W)] = \frac{1}{k-p-1} E[A(pI_m - 2H)], \tag{3.8}$$

where $H = ZW^{-1}Z'/trZW^{-1}Z'$. Let the difference in risks $\Delta_1 = EM_1(Z, W)$, where the unbiased estimator of Δ_1 is

$$\begin{aligned} M_1(Z, W) = t(Z, W)t'(Z, W) &- \frac{1}{k-p-1} \frac{b(W)}{trZW^{-1}Z'} A(pI_m - 2H) \\ &- \frac{1}{k-p-1} \frac{b(W)}{trZW^{-1}Z'} (pI_m - 2H)A. \end{aligned} \tag{3.9}$$

Note that $trST \geq 0$ for any *p.s.d.* matrices S and T and $trH = 1$. Let $T = XW_{22}^{-1}X'$. Then $A = I + T$. Using Lemma A.4 and taking trace of the second term in (3.9) yields

$$\begin{aligned} tr[G^2A(pI - 2H)] &= trG^2(pI_m - 2H) + tr[G^2T(pI_m - 2H)] \\ &= tr[G^2(pI_m - 2H)] + ptr[G^2T] - tr[T(G^2H + HG^2)] \\ &\geq g_{min}^2(pm - 2) + (p - 2\frac{g_{max}}{g_{min}})tr[G^2T] \\ &\geq (pm - 2)g_{min}^2, \end{aligned}$$

Then

$$tr[G^2M_1(Z, W)] \leq tr[t(Z, W)t'(Z, W)] - \frac{b(W)}{trZW^{-1}Z'} \cdot 2(pm - 2)g_{min}^2.$$

By the same argument in proving Theorem 2.3.2,

$$\begin{aligned} tr[G^2M_1(Y, W)] &\leq \frac{b(W)(k-p-1)^{-1}}{trZW^{-1}Z'} \left[\frac{(k-p+3)^2}{k-p-1}b(W)\lambda_{\min}^{-1}(W)tr[G^2] - 2(pm-2)\right] \\ &\leq 0 \end{aligned}$$

Q.E.D.

3.1.2 Estimators by applying MANOVA results conditionally

Now we know that

$$Z^*|X, W_{22} \sim N(\mu^*, I_m \otimes \Sigma). \tag{3.10}$$

An application of Theorem 2.2.1 to this model (3.10) conditionally gives the improved estimator $\delta_2^*(Z^*, W)$ of μ^*, consequently, the estimator of μ is given by

$$\delta_2(Z, W) = A^{1/2}\delta_2^*(Z^*, W), \tag{3.11}$$

where

$$\begin{aligned} \delta_2^*(Z^*, W) &= Z^* - t^*(Z^*, W), \\ t^*(Z^*, W) &= \frac{b(W)}{trZ^*W^{-1}Z^{*\prime}}(M^*Z^*W^{-1} + \frac{2}{k-p-1}Z^*U(W)), \end{aligned} \tag{3.12}$$

and

$$Z^* = A^{-1/2}Z, \ M^* = I + \frac{2}{k-p-1}\frac{A^{-1/2}ZW^{-1}Z'A^{-1/2}}{trA^{-1}ZW^{-1}Z'}. \tag{3.13}$$

Formally we have

Theorem 3.1.2. Suppose the regularity conditions of and condition (3.4) hold, then $\delta_2(Z,W) = Z - t_2(Z,W)$ dominates Z in risk under matrix loss where

$$t_2(Z,W) = \frac{b(W)}{trZW^{-1}Z'}(M^*ZW^{-1} + \frac{2}{n^* - p - q - 1}ZU(W)).$$

and M^* is defined by (3.13), $U(W)$ is defined by (2.6).

PROOF: Easy, since

$$\begin{aligned} &R(\delta_2;\mu,\Sigma) - R(\delta_0;\mu,\Sigma) \\ &= E^{X,W_{22}}E^{Z|X,W_{22}}((\delta_2-\mu)(\delta_2-\mu)' - (Z-\mu)(Z-\mu)') \\ &= E^{X,W_{22}}A^{\frac{1}{2}}[E^{Z^*|X,W_{22}}((\delta_2^*-\mu^*)(\delta_2^*-\mu^*)' - (Z^*-\mu^*)(Z^*-\mu^*)')]A^{\frac{1}{2}}. \end{aligned}$$

Q.E.D.

Similarly if the scalar quadratic loss L^* is adopted, we have

Corollary 3.1.2. If $pm \geq 3$ and

$$b(W) \leq 2(p - \frac{2}{m})\frac{n^* - p - q - 1}{(n^* - p - q + 3)^2}\lambda_{\min}(W),$$

then $\delta_2(Z,W)$ dominates Z in risk.

However, from Lemma 3.1.1, conditional on X and W_{22}:

$$\begin{pmatrix} Y \\ B' \end{pmatrix} \sim N\left(\begin{pmatrix} \mu + X\beta' \\ \beta' \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & W_{22}^{-1} \end{pmatrix} \otimes \Sigma\right).$$

Thus the problem is reduced to one of MANOVA. Let

$$\theta = \begin{pmatrix} \mu \\ \beta' \end{pmatrix}, \quad X^* = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix}, \quad \eta = \begin{pmatrix} \mu + X\beta' \\ \beta' \end{pmatrix} = X^*\theta.$$

Then classes of improved estimators (say, δ_J) for η can be easily obtained. Since, as is pointed in section 2.1, for any matrix Ψ, $\Psi\delta_J$ is an improved estimator of $\Psi\eta$. Let $V = \begin{pmatrix} Y \\ B' \end{pmatrix}$ and $\Psi = (\, I \quad -X\,)$, then the improved estimator for $\Psi\eta (= \mu)$ is $\Psi\delta_J$, where

$$\delta_J(V,W) = C^{1/2}\delta_C(C^{-1/2}V,W) \tag{3.14}$$

with $C = X^*$ in the $\delta_C = \delta$ defined by (2.13). We therefore obtained a different class of improved estimators for μ. Moreover, by choosing different X^* matrix we can obtain improved estimators for the slope β, in fact if let $X^* = (0, I)$ then (3.14) gives the improved estimators for β. Finally, other classes of improved estimators obtained in section 2 can also be applied to the GMANOVA model by this argument.

3.2 Comparisons of the Classes of Estimators.

Optimality is certainly what is always desired. However, when Σ is unknown, none of the improved estimators is possibly optimal (for example, admissible, even when $m = 1$, the usual normal mean estimation problem). It is hence of general interest to make comparisons among those classes of improved estimators. For example, intuitively we can expect $\delta_2(Z,W)$ to do better than $\delta_1(Z,W)$, since $\delta_2(Z,W)$ is in fact based on the conditional sufficient statistics $(Z, W_{11.2}, B, X, W_{22})$. Let the unbiased estimator of the difference in risks of $\delta_i(Z,W)$ over Z be $M_i(Z,W), i = 1,2$. Then (3.8) gives $M_1(Z,W)$, and Theorem 2.1.1 applied to (3.10) gives

$$M_2(Z,W) = A^{1/2} M^*(Z^*, W) A^{1/2},$$

where

$$M^*(Z^*,W) = t^*(Z^*,W)t^{*'}(Z^*,W) - \frac{2}{k-p-1}\frac{b(W)}{trZ^*W^{-1}Z^{*'}}(p - 2\frac{Z^{*'}W^{-1}Z^*}{trZ^*W^{-1}Z^{*'}})$$

and $t^*(Z^*,W)$ is defined by (3.12).

So

$$R(\delta_2;\mu,\Sigma) - R(\delta_1;\mu,\Sigma) = E(M_2(Z,W) - M_1(Z,W)).$$

It then suffices to show $M_2 - M_1 \le 0$. But we are only able to show this for the case when $m = 1$, that is, $Z' = z$ is a $p \times 1$ vector and $z = y - Bx$, $A = I_m + XW_{22}^{-1}X' = 1 + x'W_{22}^{-1}x$, then

$$M_1(z,W) = t'(z,W)t(z,W) - A\frac{2(p-2)}{k-p-1}\frac{b(W)}{z'W^{-1}z}$$

and

$$M_2(z,W) = A^2(t'(z,W)t(z,W) - \frac{2(p-2)}{k-p-1}\frac{b(W)}{z'W^{-1}z}),$$

where

$$t(z,W) = \frac{b(W)}{k-p-1}\frac{1}{z'W^{-1}z}((k-p+1)W^{-1}z + 2U(W)z).$$

Therefore

$$M_2(z,W) = M_1(z,W) = (A^2-1)t'(z,W)t(z,W) - (A^2-A)\frac{2(p-2)}{k-p-1}\frac{b(W)}{z'W^{-1}z}.$$

By Lemma A.3,

$$\begin{aligned} t'(z,W)t(z,W) &\le \frac{b^2(W)}{(k-p-1)^2(z'W^{-1}z)}z'W^{-2}z\{(k-p+1)+2\}^2 \\ &= \frac{b(W)}{(k-p-1)z'W^{-1}z}\frac{(k-p+3)^2}{k-p-1}b(W)\lambda_{\min}^{-1}(W). \end{aligned}$$

So

$$\begin{aligned} M_2 - M_1 &\le \frac{b(W)}{(k-p-1)}\frac{A-1}{z'W^{-1}z}[(A+1)\frac{(k-p+3)^2}{k-p-1}b(W)\lambda_{\min}^{-1}(W) - 2(p-2)A] \\ &\le 0 \end{aligned}$$

provided

$$b(W) \le \frac{A}{A+1}\frac{2(p-2)}{(k-p+3)^2}(k-p-1)\lambda_{\min}(W). \qquad (3.15)$$

But notice that $A/(A+1)$ is clearly an increasing function of $A(= 1 + x'W_{22}^{-1}x \ge 1)$, therefore we always have

$$\frac{A}{A+1} \ge \frac{1}{2}$$

Hence $b(W) \le (p-2)(k-p-1)/(k-p+3)^2\lambda_{\min}$ implies that (3.15) is true. So

$$M_2(z,W) - M_1(z,W) \le 0, \qquad \text{for all } z, W, \text{ and } A,$$

which establishes

Theorem 3.2.1. Suppose the conditions in Theorem 2.2.1 hold, $m = 1$ and if

$$b(W) \le \frac{(p-2)(k-p-1)}{(k-p+3)^2}\lambda_{\min}(W), \qquad k = n-q-1,$$

then $\delta_2(z,W)$ dominates $\delta_1(z,W)$ in risk.

Remark. When Σ is unknown, further comparisons (or admissibility) among (or of) Stein-type shrinkage estimators seem to be extremely difficult to do. However selecting among the class of improved estimators an optimal one seems to hinge upon employing prior information about the unknown parameter, which makes a Bayesian approach necessary in this regard. See Berger [1] and Berger and Haff [4] for possible Bayesian approach and some other comments.

4 APPLICATION IN SIMULATION STUDY

Control variates method has been widely used in simulation studies as a means for improving efficiency in the estimation of parameters. (See, *e.g.* , Wilson [26]). This technique collects sample data not only on the response(s) (say y) and also on some ancillary phenomenon whose true means are known, then attempts to use this extra sample information to construct an unbiased estimator of the response mean which has smaller variance than the estimator y with the same amount of simulation. For example, consider one r.v. y (a response variable for one population) with unknown mean μ which is the quantity to be estimated. Let $x = (x_1, \ldots, x_p)'$ be a r.v. with known mean $\mu_x = (\mu_1, \ldots, \mu_p)'$ and be correlated with y.

Since on each independent run of the simulation y and x result from a common probabilistic structure (e.g., a multiserver queue) and sufficient large sample is in general available, $(y, x')'$ can often be assumed to have joint normal distribution (Lavenberg and Welch [18]) with mean $(\mu_y, \mu_x')'$ and variance matrix

$$\Sigma = \begin{pmatrix} \sigma_{yy} & \sigma_{yx} \\ \sigma_{xy} & \Sigma_{xx} \end{pmatrix}$$

Then n repetitions of a simulation experiment yield statistical independent observations

$$(y_i,\ x_{1i}, x_{2i}, \ldots, x_{pi})', \qquad i = 1, 2, \ldots, n.$$

The usual regression (with random regressors) theory gives the LS (ML also) estimator of μ

$$\hat{\mu}(b) = \overline{y} - W_{yx} W_{xx}^{-1} (\overline{x} - \mu_x)$$

which is unbiased and has variance

$$\text{Var }(\hat{\mu}(b)) \leq \text{ Var }(\overline{y}) \quad \text{if } n \geq p+2 \quad \text{and } \rho^2_{y\cdot x} > \frac{p}{n-2},$$

where

$$W = \begin{pmatrix} W_{yy} & W_{yx} \\ W_{xy} & W_{xx} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n}(y_i - \overline{y})^2 & \sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})' \\ \sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x}) & \sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})' \end{pmatrix},$$

and $\rho_{y\cdot x} = \left(\sigma_{yx}\Sigma_{xx}^{-1}\sigma'_{yx}/\sigma_{yy}\right)^{1/2}$, the multiple correlation coefficient of determination.

In addition to this usual variance reduction, the combination with the shrinkage methodology offers an increased efficiency in terms of reducing the risks of the estimators for the response mean under various loss functions. Moreover, the Monte Carlo experiment in practice often involves more than one population and the responses are all vectors. (The simulation literature has largely dealt with one population case, however the multipopulation case is mentioned in Rubinstein and Marcus [24].) Suppose the populations (responses) are Y_i', $i = 1, \ldots, m$ and Y_i is a $p \times 1$ random vector, the control variates used respectively for the responses are $X_1, \ldots, X_m$, which are $q \times 1$ random vectors with known mean vectors (say) $0, \ldots, 0$. Further assume the joint distribution is

$$(Y, X) \sim N((\mu, 0), I \otimes \Sigma),$$

where

$$Y = (Y_1', \ldots, Y_m')', \qquad X = (X_1', \ldots, X_m')',$$
$$\mu = (\mu_1', \ldots, \mu_m')', \qquad 0 = (0, \ldots, 0).$$

Suppose we have (say) n independent runs of the simulation experiment, namely,

$$Y_i^{(j)}, \quad X_i^{(j)} \qquad i = 1, \ldots, m, \quad j = 1, \ldots, n.$$

and $(\overline{Y}, \overline{X})$ with

$$\overline{Y} = (\overline{Y}_1', \ldots, \overline{Y}_m')', \quad \overline{Y}_i = \frac{1}{n}\sum_{j=1}^{n} Y_i^{(j)},$$
$$\overline{X} = (\overline{X}_1', \ldots, \overline{X}_m')', \quad \overline{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_i^{(j)},$$

and

$$W_1 = \sum_{i=1}^{m} \begin{pmatrix} \sum_{j=1}^{n}(Y_i^{(j)} - \overline{Y}_i)(Y_i^{(j)} - \overline{Y}_i)' & \sum_{j=1}^{n}(Y_i^{(j)} - \overline{Y}_i)(X_i^{(j)} - \overline{X}_i)' \\ \sum_{j=1}^{n}(X_i^{(j)} - \overline{X}_i)(Y_i^{(j)} - \overline{Y}_i)' & \sum_{j=1}^{n}(X_i^{(j)} - \overline{X}_i)(X_i^{(j)} - \overline{X}_i)' \end{pmatrix}$$

$$= \begin{pmatrix} W_{yy} & W_{yx} \\ W_{xy} & W_{xx} \end{pmatrix}$$

then

$$W_1 \sim \mathcal{W}_{p+q}(\frac{1}{n}\Sigma,\ m(n-1)),$$

where

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix},$$

The usual control variates technique leads to the use of $Z_1 = \overline{Y} - \overline{X}B_1'$ and $B_1 = W_{yx}W_{xx}^{-1}$ in order to attain maximum variance reduction if the correlation between Y and X are reasonably high. This can be achieved in practice by choosing appropriate control variates. Note that

$$\text{Cov }(Z|X, W_{xx}) = (I + XW_{xx}^{-1}X') \otimes \Sigma$$

and

$$EZ = \mu, \quad \text{a constant matrix.}$$

Then

$$\begin{aligned} \text{Cov }(Z) &= E^{X,W_{xx}}E \text{ Cov }(Z|X, W_{xx}) \\ &= E^{X,W_{xx}}(I_q + XW_{xx}^{-1}X') \otimes \Sigma_{yy\cdot x} \\ &= E(I_q + \frac{q}{d-q-1}I_q) \otimes \Sigma_{yy\cdot x} \\ &= \left(\frac{d-1}{d-q-1}I_q\right) \otimes \Sigma_{yy\cdot x}, \end{aligned}$$

since $E\ x_i'W_{xx}^{-1}x_i = q/(d-q-1)$ for any $i = 1, 2, \ldots, m$, where $d = m(n-1)$, the degree of freedom of W. Straight algebra then gives that

$$\text{Cov }(Z) \leq \text{ Cov }(Y),$$

provided

$$\Sigma_{yy}^{-1/2}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1/2} \geq \frac{q}{d-1}I_p.$$

In other words we need the canonical correlations to be all at least as large as $p(n-1)^{-1}I$. Moreover, a direct application of the results in section 3 would further improve the estimator Z_1, since

$$(\overline{Y}, \overline{X}) \sim N((\mu, 0), \ \frac{1}{n}\Sigma),$$

and

$$W_1 \sim W_{p+q}(m(n-1), \ \frac{1}{n}\Sigma).$$

And then the resulting improved estimators can be easily obtained through (3.11)–(3.13), which in fact is given by

$$\delta_1(Z, W_1) = Z_1 - t(Z_1, W_1)$$

where

$$t(Z_1, W_1) = \frac{b(W_1)}{trZ_1W^{-1}Z_1'}(Z_1W_1^{-1}M_1 + \frac{2}{d-q-p-1}Z_1U_1(W_1)),$$

$$M_1 = I + \frac{2}{d-q-p-1}\frac{Z_1'Z_1'W_1^{-1}}{trZ_1W^{-1}Z_1'},$$

and

$$U(W_1) = \left(\frac{1+\delta_{ij}}{2}\frac{\partial}{\partial w_{ij}}\log b(W_1)\right).$$

APPENDIX

A.1. The first lemma concerns a well-known integration by parts identity due to Stein [25].

Lemma A.1 (Integration by Parts). Let $x \sim N_p(\theta, \Sigma)$, $\gamma = (\gamma_1(x), \ldots, \gamma_p(x))'$, and $\gamma(x)$ satisfy the regularity conditions for integration by parts. (See pp. 362 in Berger [2]). Then we have

$$E(x-\theta)'\gamma(x) = Etr\Sigma J_{\gamma(x)}(x),$$

where $J_{\gamma(x)}(x) = (\frac{\partial}{\partial x_i}\gamma_j(x))$, provided the integral at the right side exists.

A.2. Lemma A.2 (Wishart Identity). Let $W \sim \mathcal{W}_p(\Sigma, n)$ and $T(W) = (t_{ij}(W))$ satisfy the conditions on $T(W)$ specified in Haff [13]. Then we have

$$Etr[W^{-1}T(W)] = \frac{1}{n-p-1}(Etr[\Sigma^{-1}T(W)] - 2Etr[D_{(1/2)}T(W)])$$

provided the expectations exist, where

$$D_{(1/2)} = \left(\frac{1}{2-\delta_{ij}}\frac{\partial}{\partial w_{ij}}\right), \quad \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j, \end{cases}$$

and consequently, $tr[D_{(1/2)}T(W)] = \sum_{i=1}^{p} \frac{\partial}{\partial w_{ii}} t_{ii} + \frac{1}{2}\sum_{i\neq j} \frac{\partial t_{ij}}{\partial w_{ij}}$.

PROOF: Equation (2.1) in Haff [13] with $h(\cdot) = 1$.

A.3. Lemma A.3. For any two $p \times 1$ vectors x, $y (x \neq 0)$ and any two scalars c_1, c_2,

$$||x||^2(|c_1| - |c_2|\frac{||y||}{||x||})^2 \leq ||c_1 x + c_2 y||^2 \leq ||x||^2(|c_1| + |c_2|\frac{||y||}{||x||})^2,$$

where $||\cdot||$ is the usual Euclidean norm in R^p.

Proof. See for example Gleser [10].

A.4. Lemma A.4. Let H, F be $m \times m$ symmetric matrices. If $0 \leq H \leq I$ and $F > 0$, then $HF + FH \leq 2aF$, in the sense of semi-definiteness of matrices, where $a = (\lambda_{max}(F)/\lambda_{min}(F))^{1/2}$ is the condition number of F.

PROOF: Since $F \geq 0$, there exists $G \geq 0$ such that $F = G^2$. It then suffices to prove that $G^{-1}HG + GHG^{-1} \leq aI$. For any matrix A, let $\lambda(A)$ denote any one of its eigenvalues. Since $\lambda((aI - G^{-1}HG + GHG^{-1})/2) = a - \lambda((G^{-1}HG + GHG^{-1})/2)$, it suffices to show that $\lambda((G^{-1}HG + GHG^{-1})/2) \leq a$. By the Singular-Value Decomposition, there exists orthogonal matrices U and V such that $G^{-1}HG = UDV$ where

$$D = Diag(\lambda_1^{1/2}(HFHF^{-1}), \cdots, \lambda_m^{1/2}(HFHF^{-1}),$$

and $\lambda_i(A)$ denotes the i^{th} eigenvalue of the matrix A. Let $Q = UV$ and $P = V'DV$, we then have $G^{-1}HG = QP$, where the matrix Q is orthogonal and P is positive semi-definite. (This is known as the polar-decomposition of a matrix, see for example Marshall and Olkin [20], pp 498-501). Then for any vector a such that $(a'a)^{1/2} = 1$, we have

$$a'(\frac{G^{-1}HG + GHG^{-1}}{2})a = a'QPa \leq (a'QQ'a)^{1/2}(a'P'Pa)^{1/2} = (a'P'Pa)^{1/2}.$$

Consequently, the classic Courant and Fischer representation of eigenvalues gives

$$\lambda(\frac{G^{-1}HG+GHG^{-1}}{2}) \le \lambda^{1/2}(P'P) = \lambda^{1/2}(HFHF^{-1})$$

However (see for example Marshall and Olkin [20], pp. 246 - 248),

$$\lambda^2(HFHF^{-1}) \le \lambda_{max}(FH^2F)\lambda_{max}(F^{-1}H^2F^{-1}).$$

But $0 \le H \le I$ implies $H^2 \le I$ and $F > 0$ implies $\lambda^{1/2}(F^2) = \lambda(F)$, then

$$\lambda(HFHF^{-1}) \le \lambda_{max}(F)\lambda_{min}^{-1}(F).$$

Therefore

$$\lambda(\frac{G^{-1}HG+GH^{-1}}{2}) \le \left(\frac{\lambda_{max}}{\lambda_{min}}\right)^{1/2} = a.$$

Q.E.D.

A.5. (PROOF of Equation (3.8) in section 3.1.1.) Since $A > 0$, $A = S'S = S^2$ where S is symmetric and $S > 0$. Then from (3.7) we have

$$\begin{aligned} E(Z-\mu)t'(Z,W) &= \frac{1}{k-p-1}\ E\ (Z-\mu)(h(Z,W)W\Sigma^{-1})' \\ &= \frac{1}{k-p-1}\ S\ E\ (Z^*-\mu^*)(h(SZ^*,W)W\Sigma^{-1})', \end{aligned}$$

where

$$Wh_j(SZ^*,W) = \frac{b(W)}{trSZ^*W^{-1}Z^*S}(SZ^*)'_j.$$

Noting that $Z_i^* \sim N(\mu^*, \Sigma)$ and applying Lemma A.1 to the (i,j) element of the integral yields that

$$\begin{aligned} E\ (Z_i^*-\mu_i^*)'\Sigma^{-1}Wh_j(SZ^*,W) &= E\ \left[tr[J_{Wh_j(SZ^*,W)}(Z_i^*)]\right] \\ &= E\ \left[tr\left(\sum_{k=1}^{m}\frac{\partial}{\partial Z_k}\left(\frac{b(W)}{trZW^{-1}Z'}Z_j\right)\frac{\partial Z_k}{\partial Z_i^*}\right)\right] \end{aligned}$$

Straightforward calculus gives that

$$\frac{\partial}{\partial Z_k}\frac{b(W)}{trZW^{-1}Z'}Z_j = \frac{\delta_{kj}I_p}{trZW^{-1}Z'} - 2\frac{W^{-1}Z_kZ_j'}{(trZW^{-1}Z')^2},$$

where δ_{kj} is the Kronecker delta. Also note that $Z = SZ^*$ and $Z_k = Z^{*'} S e_k$ and taking derivative directly yields

$$\frac{\partial Z_k}{\partial Z_i^*} = s_{ik} I_p.$$

Then the (i, j) element of $T_R[W \circ J(Z, W)]$ is

$$tr[J_{(Wh_j)}(Z_i)] = \frac{b(W)}{trZW^{-1}Z'}(s_{ij}p - 2\frac{(SZW^{-1}Z')_{ij})}{trZW^{-1}Z'}).$$

Then

$$E\ (Z^* - \mu^*)(h(SZ^*, W)W\Sigma^{-1})' = S\frac{b(W)}{trZW^{-1}Z'}(pI_m - \frac{ZW^{-1}Z'}{trZW^{-1}Z'}).$$

$$E(Z - \mu)t'(Z, W) = \frac{1}{k - p - 1}\ E\ \frac{b(W)}{trZW^{-1}Z'}A(pI_m - 2\frac{ZW^{-1}Z'}{trZW^{-1}Z'}).$$

Q.E.D.

Acknowledgements. The author wishes to express his deep gratitude to his major Professor, Leon J. Gleser for his invaluable help in the preparation of this manuscript.

REFERENCES

[1] Berger, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8**, 716–761.

[2] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer–Verlag, New York.

[3] Berger, J., Bock, M.E., Brown, L.D., Casella, G. and Gleser, L.J. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* **5**, 763–771.

[4] Berger, J. and Haff, L.R. (1983). A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Statistics and Decisions* **1**, 105–129.

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

[6] Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations – An extension of Stein's method. *Biometrika* **59**, 335–347.

[7] Brown, P.J. and Zidek, J.V. (1980). Adaptive multivariate ridge regression. *Ann. Statist.* , **8**, 64 – 74.

[8] Geisser, S. (1970). Bayesian analysis of growth curve. *Sankhyá* **A32**, 53–64.

[9] Gleser, L. (1979). Minimax estimation of a normal mean vector when the covariance matrix is unknown. *Ann. Statist.* **7**, 838 – 846.

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

[11] Gleser, L. (1988). New estimators for the mean vector of a normal distribution with unknown covariance matrix. *In Statistical Decision Theory and Related Topics IV*, **Vol. 1**, (S.S. Gupta and J.O. Berger, eds). 347-360 Springer–Verlag, New York.

[12] Gleser, L. and Olkin, I. (1970). Linear models in multivariate analysis. *In Essays in Probability and Statistics* (R.C. Bose, I.M. Chakravarti, P.C. Mahalanobis, C.R. Rao and K.J.C. Smith, eds.). 267-292 University of North Carolina Press: Chapel Hill, N.C.

[13] Haff, L. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.* **9**, 531–542.

[14] Honda, T. (1990). Unpublished manuscript.

[15] Hooper, P. (1982). Invariant confidence sets with smallest expected measure. *Ann. Statist.* **10**, 1283–1294.

[16] Kariya, T. (1985). *Testing in the Multivariate General Linear Model.* Kinokaniya Co. Ltd: Tokyo.

[17] Kubokawa, T. and Saleh, A.K.Md. (1990). Unpublished manuscript.

[18] Lavenberg, S.S. and Welch, P.D. (1980). A perspective on the use of control variates to increase the efficiency of Monte Carlo simulation. *Management Sci*, **27**, 332–335.

[19] Lin, P.E. and Tsai, H.L. (1973). Generalized Bayes minimax estimators [f the multivariate normal mean with with unknown covariance matrix. *Ann. Statist.* **1**, 142-145.

[20] Marshall, A.W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its applications.* Academic Press, New York.

[21] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley: New York.

[22] Potthoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–326.

[23] Rao, C.R. (1965). The theory of least squares when the parameters are stochastic and its applications to the analysis of growth curves. *Biometrika* **52**, 447–458.

[24] Rubinstein, R. and Marcus, R. (1985). Efficiency of multivariate control variates in Monte Carlo simulation. *Operations Res.* **33**, 661–667.

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

[26] Wilson, J.R. (1984). Variance reduction techniques for digital simulation. *Math. and Man. Sci.* **1**, 227–312.

[27] Zheng, Z. (1986). On estimation of matrix of normal mean. *J. Multivariate Anal.* **18**, 70–82.

[28] Zidek, J. (1978). Deriving unbiased risk estimators of multinormal mean and regression coefficient estimators using zonal polynomials. *Ann. Statist.* **6**, 769 – 782.