PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 36, Fasc. 2 (2016), pp. 187–200

MINIMAX ESTIMATION OF THE MEAN MATRIX OF THE MATRIX-VARIATE NORMAL DISTRIBUTION

BY

S. ZINODINY (TEHRAN), S. REZAEI (TEHRAN),
AND S. NADARAJAH (MANCHESTER)

Abstract. In this paper, the problem of estimating the mean matrix $\boldsymbol{\Theta}$ of a matrix-variate normal distribution with the covariance matrix $\mathbf{V} \otimes \mathbf{I}_m$ is considered under the loss functions, $\omega \operatorname{tr}((\boldsymbol{\delta} - \mathbf{X})'\mathbf{Q}(\boldsymbol{\delta} - \mathbf{X})) + (1-\omega) \operatorname{tr}((\boldsymbol{\delta} - \boldsymbol{\Theta})'\mathbf{Q}(\boldsymbol{\delta} - \boldsymbol{\Theta}))$ and $k[1 - e^{-\operatorname{tr}((\boldsymbol{\delta}-\boldsymbol{\Theta})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{\delta}-\boldsymbol{\Theta}))}]$. We construct a class of empirical Bayes estimators which are better than the maximum likelihood estimator under the first loss function for $m > p+1$ and hence show that the maximum likelihood estimator is inadmissible. For the case $\mathbf{Q} = \mathbf{V} = \mathbf{I}_p$, we find a general class of minimax estimators. Also we give a class of estimators that improve on the maximum likelihood estimator under the second loss function for $m > p+1$ and hence show that the maximum likelihood estimator is inadmissible.

2010 AMS Mathematics Subject Classification: Primary 62C15; Secondary 62C20.

Key words and phrases: Empirical Bayes estimation, matrix-variate normal distribution, mean matrix, minimax estimation.

1. INTRODUCTION

Let $\mathbf{X} = (x_{i,j})$ be a $p \times m$ random matrix having a matrix-variate normal distribution with mean matrix $\boldsymbol{\Theta} = (\theta_{i,j})$ and covariance matrix $\mathbf{V} \otimes \mathbf{I}_m$, where $\mathbf{V}$ is known, $\mathbf{I}_k$ is the $k \times k$ identity matrix, and $\otimes$ indicates the Kronecker product. This note considers estimation of $\boldsymbol{\Theta}$ relative to the loss functions

$$L_1(\boldsymbol{\delta}; \boldsymbol{\Theta}) = \omega \operatorname{tr}\big((\boldsymbol{\delta} - \mathbf{X})'\mathbf{Q}(\boldsymbol{\delta} - \mathbf{X})\big) + (1-\omega) \operatorname{tr}\big((\boldsymbol{\delta} - \boldsymbol{\Theta})'\mathbf{Q}(\boldsymbol{\delta} - \boldsymbol{\Theta})\big) \tag{1.1}$$

and

$$L_2(\boldsymbol{\delta}; \boldsymbol{\Theta}) = k\big[1 - e^{-\operatorname{tr}\left((\boldsymbol{\delta}-\boldsymbol{\Theta})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{\delta}-\boldsymbol{\Theta})\right)}\big], \tag{1.2}$$

where $k > 0$, $0 < \omega < 1$, $\boldsymbol{\Gamma}$ and $\mathbf{Q}$ are known $p \times p$ positive definite matrices, and $\operatorname{tr}(\mathbf{A})$ and $\mathbf{A}'$ denote, respectively, the trace and the transpose of a matrix $\mathbf{A}$. The loss function (1.1) is an extension of the loss function introduced by Asgharazadeh and Farsipour [1] to the multivariate case. It is also a multivariate extension of Zellner's [16] balanced loss function. See also Dey et al. [3]. It is formulated to reflect two criteria, namely goodness of fit and precision of estimation. The loss function (1.2) is an extension of the extended reflected normal loss function to the multivariate case. It was introduced by Spiring [12] for the first time.

The balanced loss function takes both error of estimation and goodness of fit into account. The classical loss function considers only error of estimation. The problem of estimating the mean matrix of matrix-variate normal distributions under balanced loss functions is important in multiple regression models. As pointed out in Leung and Spiring [9], estimation under the loss function (1.2) is important for the determination of optimal operating conditions for a process, the determination of average loss per unit produced and monitoring of loss associated with a process. The loss function (1.2) is also important for evaluating manufacturing and environmental risks (Pan [11]) and bank loan approval decisions (Majeske and Lauer [10]).

Efron and Morris [4] extended the so-called Stein effect and proposed an empirical Bayes estimator outperforming the maximum likelihood estimator, $\mathbf{X}$, for the case $m > p + 1$. Since then many classes of minimax estimators better than the maximum likelihood estimator have been found for quadratic loss and general quadratic loss functions, see Stein [13], Zhang [17], [18], Ghosh and Shieh [5] and Tsukuma [15]. However, little is known about improving on the maximum likelihood estimator for other non-quadratic loss functions. This note considers this problem under the given loss functions for the matrix-variate normal distribution.

The contents of this note are organized as follows. In Section 2, we obtain a class of empirical Bayes estimators dominating the maximum likelihood estimator under the loss function (1.1) for $m > p + 1$, showing that the maximum likelihood estimator is inadmissible. Also, when $\mathbf{Q} = \mathbf{V} = \mathbf{I}_p$, we find a general class of minimax estimators using a technique of Stein [13]. In fact, we extend the results of Asgharazadeh and Farsipour [1] to the multivariate case. In Section 3, we find a class of minimax estimators better than the maximum likelihood estimator under the loss function (1.2) for $m > p + 1$ and extend the results of Towhidi and Behboodian [14] to the multivariate case. The minimax estimators of Sections 2 and 3 are compared by simulation in Section 4.

2. MINIMAX ESTIMATION OF THE MEAN MATRIX UNDER THE LOSS FUNCTION (1.1)

Assume

$$\mathbf{X} \sim N_{p\times m}(\mathbf{\Theta}, \mathbf{V} \otimes \mathbf{I}_m),$$

where $N_{p\times m}(\mathbf{\Theta}, \mathbf{V} \otimes \mathbf{I}_m)$ stands for the matrix-variate normal distribution with mean matrix $\mathbf{\Theta}$ and covariance matrix $\mathbf{V} \otimes \mathbf{I}_m$. Here, we consider the problem of estimating $\mathbf{\Theta}$ under the loss function (1.1). A natural candidate for estimating $\mathbf{\Theta}$ is the maximum likelihood estimator $\mathbf{X}$.

Lemma 2.1 shows that the maximum likelihood estimator is minimax under the loss function (1.1).

LEMMA 2.1. *The estimator* $\mathbf{X}$ *is minimax under the loss function* (1.1).

P r o o f. Note that the risk function of $\mathbf{X}$ is

$$R_1(\mathbf{X}, \mathbf{\Theta}) = E_{\mathbf{\Theta}}\left[L_1(\mathbf{X}; \mathbf{\Theta})\right] = (1-\omega)m\,\mathrm{tr}(\mathbf{QV}).$$

Now, suppose Θ has a matrix-variate normal distribution with zero mean matrix and a covariance matrix $n\mathbf{C}\otimes\mathbf{I}_m$, i.e., $\mathbf{\Theta}\sim N_{p\times m}(\mathbf{0}, n\mathbf{C}\otimes\mathbf{I}_m)$, where $\mathbf{C}$ is a $p\times p$ known positive definite matrix. Then the posterior distribution is

$$\mathbf{\Theta}|\mathbf{X}\sim N_{p\times m}\left(\left(\mathbf{I}_p-\left(\mathbf{V}^{-1}+\frac{\mathbf{C}^{-1}}{n}\right)^{-1}\frac{\mathbf{C}^{-1}}{n}\right)\mathbf{X}, \left(\mathbf{V}^{-1}+\frac{\mathbf{C}^{-1}}{n}\right)^{-1}\otimes\mathbf{I}_m\right),$$

and the marginal distribution is

$$\mathbf{X}\sim N_{p\times m}\left(\mathbf{0}, n\mathbf{C}\left(\mathbf{V}^{-1}+\frac{\mathbf{C}^{-1}}{n}\right)^{-1}\mathbf{V}\otimes\mathbf{I}_m\right).$$

Thus, the Bayes estimator is

$$\omega\mathbf{X}+(1-\omega)E_{\mathbf{\Theta}}\left[\mathbf{\Theta}|\,\mathbf{X}\right]=\left(\mathbf{I}_p-(1-\omega)\left(\mathbf{V}^{-1}+\frac{\mathbf{C}^{-1}}{n}\right)^{-1}\frac{\mathbf{C}^{-1}}{n}\right)\mathbf{X}.$$

The corresponding Bayes risk is

$$\begin{aligned} r_n &= (1-\omega)m\,\mathrm{tr}\left(\mathbf{Q}\left(\mathbf{V}^{-1}+\frac{\mathbf{C}^{-1}}{n}\right)^{-1}\right) \\ &\quad + m\omega(1-\omega)\,\mathrm{tr}\left(\mathbf{QV}\frac{\mathbf{C}^{-1}}{n}\left(\mathbf{V}^{-1}+\frac{\mathbf{C}^{-1}}{n}\right)^{-1}\right). \end{aligned}$$

Since $\mathbf{V}$ and $\mathbf{C}$ are positive definite matrices, there exist a nonsingular matrix $\mathbf{T}$ and diagonal matrix $\mathbf{\Lambda}=\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ such that

$$\mathbf{C}^{-1}=\mathbf{T\Lambda T}', \quad \mathbf{V}^{-1}=\mathbf{TT}',$$

so

$$r_n=(1-\omega)m\,\mathrm{tr}\left(\mathbf{Z}\left(\mathbf{I}_p+\frac{\mathbf{\Lambda}}{n}\right)^{-1}\right)+m\omega(1-\omega)\,\mathrm{tr}\left(\mathbf{Z}\frac{\mathbf{\Lambda}}{n}\left(\mathbf{I}_p+\frac{\mathbf{\Lambda}}{n}\right)^{-1}\right),$$

where $\mathbf{Z} = \mathbf{T}^{-1}\mathbf{Q}\left(\mathbf{T}'\right)^{-1}$. If $z_{i,j}$ denotes the (i,j)th element of $\mathbf{Z}$, then

$$r_n = (1-\omega)m\sum_{i=1}^{p}\frac{nz_{i,i}}{n+\lambda_i} + m\omega(1-\omega)\sum_{i=1}^{p}\frac{\lambda_i z_{i,i}}{n+\lambda_i},$$

and

$$\begin{aligned}\lim_{n\to\infty} r_n &= (1-\omega)m\sum_{i=1}^{p} z_{i,i} = (1-\omega)m\,\mathrm{tr}(\mathbf{Z})\\ &= (1-\omega)m\,\mathrm{tr}(\mathbf{QV}) = \sup_{\boldsymbol{\Theta}} R(\mathbf{X};\boldsymbol{\Theta}).\end{aligned}$$

Hence, by Theorem 1.12 of Lehmann and Casella [8], p. 316, and the results of Blyth [2], $\mathbf{X}$ is minimax under the loss function (1.1). ∎

We now construct a class of empirical Bayes estimators better than $\mathbf{X}$.

Assume that we have some additional information about $\boldsymbol{\Theta}$ that can be written as follows:

$$\boldsymbol{\Theta} \sim N_{p\times m}(\mathbf{0}_{p\times m}, \mathbf{A}\otimes\mathbf{I}_m).$$

The conditional distribution of $\boldsymbol{\Theta}$ given $\mathbf{X}$ is

$$N_{p\times m}\Big(\big(\mathbf{I}_p - \mathbf{V}(\mathbf{V}+\mathbf{A})^{-1}\big)\mathbf{X}, (\mathbf{V}^{-1}+\mathbf{A}^{-1})^{-1}\otimes\mathbf{I}_m\Big),$$

so the Bayes estimator of $\boldsymbol{\Theta}$ under the loss function (1.1) is

$$\begin{aligned}\boldsymbol{\delta}_{B,\omega}(\mathbf{X}) &= \omega\mathbf{X} + (1-\omega)E\left[\boldsymbol{\Theta}|\mathbf{X}\right] = \omega\mathbf{X} + (1-\omega)(\mathbf{I}_p - \mathbf{V}\boldsymbol{\Sigma}^{-1})\mathbf{X}\\ &= \left(\mathbf{I}_p - (1-\boldsymbol{\omega})\mathbf{V}\boldsymbol{\Sigma}^{-1}\right)\mathbf{X},\end{aligned}$$

where $\boldsymbol{\Sigma} = \mathbf{A}+\mathbf{V}$. In an empirical Bayes scenario, $\boldsymbol{\Sigma}$ is unknown and is estimated from the marginal distribution of $\mathbf{X}$. The marginal distribution of $\mathbf{X}$ is $N_{p\times m}\left(\mathbf{0}_{p\times m}, \boldsymbol{\Sigma}\otimes\mathbf{I}_m\right)$. Hence, $\mathbf{S} = \mathbf{XX}'$ is completely sufficient for $\boldsymbol{\Sigma}$. Therefore, an empirical Bayes estimator is

$$\boldsymbol{\delta}_{EB,\omega}(\mathbf{X}) = \big(\mathbf{I}_p - (1-\omega)\mathbf{V}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\big)\mathbf{X}, \tag{2.1}$$

where $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ is an estimator for $\boldsymbol{\Sigma}^{-1}$ and $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ depends on $\mathbf{X}$ only through $\mathbf{S}$. According to Ghosh and Shieh [5], a natural candidate for the estimator $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ is $(m-p-1)\mathbf{S}^{-1}$. Ghosh and Shieh [5] obtained the empirical Bayes estimator

$$\boldsymbol{\delta}_{EB}(\mathbf{X}) = \big(\mathbf{I}_p - \mathbf{V}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\big)\mathbf{X}$$

when the loss function was

$$L_3\left(\boldsymbol{\delta};\boldsymbol{\Theta}\right) = \mathrm{tr}\left((\boldsymbol{\delta}-\boldsymbol{\Theta})'\mathbf{Q}(\boldsymbol{\delta}-\boldsymbol{\Theta})\right). \tag{2.2}$$

Theorem 2.1 below considers the relationship between $\boldsymbol{\delta}_{EB}(\mathbf{X})$ and $\boldsymbol{\delta}_{EB,\omega}(\mathbf{X})$. To continue discussion, we need the following notation borrowed from Ghosh and Shieh [5]: let $E_{\boldsymbol{\Theta}}$ denote the expectation conditional on $\boldsymbol{\Theta}$, $\widetilde{E}$ the expectation over the marginal distribution of $\boldsymbol{\Theta}$, E the expectation over the joint distribution of $\mathbf{X}$ and $\boldsymbol{\Theta}$. Let $R_i(\boldsymbol{\delta};\boldsymbol{\Theta}) = E_{\boldsymbol{\Theta}}[L_i(\boldsymbol{\delta};\boldsymbol{\Theta})]$, $i=1,2,3$, and $\mathbf{D}=(d_{i,j})$, where

$$d_{i,j} = \left(\frac{1+\delta_{i,j}}{2}\right)\frac{\partial}{\partial s_{i,j}}$$

(see Ghosh and Shieh [5], the last line of p. 308), $\delta_{i,j}$ being the Kronecker deltas, and $s_{i,j}$ the (i,j)th element of $\mathbf{S}$. Note that $\mathbf{D}$ is a differential operator in the form of a matrix. If we assume that $f(\mathbf{A})$ is a real-valued function of a $p\times p$ matrix $\mathbf{A}$, then $\mathbf{D}f(\mathbf{A})$ is a $p\times p$ matrix with (i,j)th element

$$\frac{1+\delta_{i,j}}{2}\frac{\partial f(\mathbf{A})}{\partial s_{i,j}}.$$

For example, if $p=2$ and $f(\mathbf{A}) = s_{1,1}^2 + s_{2,2}^2$, then

$$\mathbf{D}f(\mathbf{A}) = \begin{pmatrix} 2s_{1,1} & 0 \\ 0 & 2s_{2,2} \end{pmatrix}.$$

Also, for any $p\times p$ matrix $\mathbf{T}$, it follows that $\mathbf{D}(\mathbf{T})$ is a $p\times p$ matrix with (i,j)th element $\sum_{l=1}^{p} d_{i,l}t_{l,j}$. For example, if $p=2$, then

$$\mathbf{D}(\mathbf{S}) = \begin{pmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$

THEOREM 2.1. *The empirical Bayes estimator* (2.1) *is minimax under the loss function* (1.1) *if $\boldsymbol{\delta}_{EB}(\mathbf{X})$ is minimax under the quadratic loss function* (2.2). *Also, the risk function of $\boldsymbol{\delta}_{EB,\omega}(\mathbf{X})$ under the loss function* (1.1) *is*

$$(2.3)\quad R_1(\boldsymbol{\delta}_{EB,\omega};\boldsymbol{\Theta}) = R_1(\mathbf{X};\boldsymbol{\Theta}) + (1-\omega)^2 E\Big[\operatorname{tr}\big(\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{Q}_0\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{S}\big) - 4\operatorname{tr}\Big(\mathbf{D}\big(\mathbf{S}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{Q}_0\big)\Big) - 2(m-p-1)\operatorname{tr}\big(\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{Q}_0\big)\Big],$$

where $\mathbf{Q}_0 = \mathbf{VQV}$.

Proof. We will prove this theorem using a technique applied in Ghosh and Shieh [5]. We have

$$E[L_1(\boldsymbol{\delta}_{EB,\omega};\boldsymbol{\Theta})] = E\big[\sum_{k=1}^{m}\omega(\boldsymbol{\delta}_{EB,\omega}^{k}-\mathbf{X}_k)'\mathbf{Q}(\boldsymbol{\delta}_{EB,\omega}^{k}-\mathbf{X}_k) + (1-\omega)(\boldsymbol{\delta}_{EB,\omega}^{k}-\boldsymbol{\theta}_k)'\mathbf{Q}(\boldsymbol{\delta}_{EB,\omega}^{k}-\boldsymbol{\theta}_k)\big],$$

where $\boldsymbol{\delta}_{EB,\omega}(\mathbf{X}) = (\boldsymbol{\delta}^1_{EB,\omega}, \boldsymbol{\delta}^2_{EB,\omega}, \ldots, \boldsymbol{\delta}^m_{EB,\omega})$ and $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)$, with $\boldsymbol{\delta}^k_{EB,\omega}$ and $\boldsymbol{\theta}_k$ being p-variate vectors. Since the conditional distribution of $\boldsymbol{\Theta}$ given $\mathbf{X}$ is $N_{p\times m}\big(\boldsymbol{\delta}_{B,\omega}, (\mathbf{V}^{-1} + \mathbf{A}^{-1})^{-1} \otimes \mathbf{I}_m\big)$, we have

$$E\left[L_1\left(\boldsymbol{\delta}_{EB,\omega};\boldsymbol{\Theta}\right)\right] = E\left[L_1\left(\boldsymbol{\delta}_{B,\omega};\boldsymbol{\Theta}\right)\right] + (1-\omega)^2\, E\Big[\operatorname{tr}\Big(\big(\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S}) - \boldsymbol{\Sigma}^{-1}\big)\mathbf{Q}_0\big(\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S}) - \boldsymbol{\Sigma}^{-1}\big)\mathbf{S}\Big)\Big].$$

Applying the Wishart identity (Haff [6]) and $E\left(\mathbf{S}\right) = m\boldsymbol{\Sigma}$, we obtain

$$\begin{aligned}(2.4)\quad &\widetilde{E}\left[L_1\left(\boldsymbol{\delta}_{EB,\omega};\boldsymbol{\Theta}\right)\right] \\ &= m\,(1-\omega)\operatorname{tr}(\mathbf{QV}) + (1-\omega)^2\,\widetilde{E}\Big[E_{\boldsymbol{\Theta}}\Big[\operatorname{tr}\big(\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{Q}_0\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{S}\big) \\ &\quad - 4\operatorname{tr}\Big(\mathbf{D}\big(\mathbf{S}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{Q}_0\big)\Big) - 2(m-p-1)\operatorname{tr}\big(\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{Q}_0\big)\Big]\Big].\end{aligned}$$

The conditional distribution of $\mathbf{S} = \mathbf{XX}'$ given $\boldsymbol{\Theta}$ depends on $\boldsymbol{\Theta}$ only through $\boldsymbol{\Theta\Theta}' = \boldsymbol{\Lambda}$, say. The family of distributions of $\boldsymbol{\Lambda}$ being Wishart $W_p\left(m, \boldsymbol{\Lambda}\right)$ is complete for $\mathbf{A}$ or $\boldsymbol{\Sigma}$. Also $R_1\left(\mathbf{X};\boldsymbol{\Theta}\right) = k\left(1-\omega\right)\operatorname{tr}(\mathbf{QV})$ for all $\boldsymbol{\Theta}$ and

$$R_1\left(\boldsymbol{\delta}_{EB,\omega};\boldsymbol{\Theta}\right) - R_1\left(\mathbf{X};\boldsymbol{\Theta}\right) = (1-\omega)^2\,E\left[R_3\left(\boldsymbol{\delta}_{EB};\boldsymbol{\Theta}\right) - R_3\left(\mathbf{X};\boldsymbol{\Theta}\right)\right].$$

Hence, if $R_3\big(\boldsymbol{\delta}_{EB}(\mathbf{X});\boldsymbol{\Theta}\big) \leqslant R_3\left(\mathbf{X};\boldsymbol{\Theta}\right)$, then $R_1\big(\boldsymbol{\delta}_{EB,\omega}(\mathbf{X});\boldsymbol{\Theta}\big) \leqslant R_1\left(\mathbf{X};\boldsymbol{\Theta}\right)$, and (2.3) follows immediately from (2.4). The proof is complete. ∎

If $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ is chosen in such a way that $\boldsymbol{\delta}_{EB}$ improves on $\mathbf{X}$ under the quadratic loss function (2.2), then $\boldsymbol{\delta}_{EB,\omega}$ improves on $\mathbf{X}$ also under the loss function (1.1). Let $\mathbf{O}_p$ be the set of orthogonal matrices of order p and let $\mathbf{V}_{m,p}$ be the Stiefel manifold, namely, $\mathbf{V}_{m,p} = \{\mathbf{V} \in \mathbf{R}^{m\times p}, \mathbf{V}'\mathbf{V} = \mathbf{I}_p\}$. Write the singular value decomposition of $\mathbf{X}$ as $\mathbf{UL}^{1/2}\mathbf{V}'$, where $\mathbf{U} \in \mathbf{O}_p$, $\mathbf{V} \in \mathbf{V}_{m,p}$ and $\mathbf{L} = \operatorname{diag}(l_1, l_2, \ldots, l_p)$ with $l_1 > l_2 > \ldots > l_p > 0$. Ghosh and Shieh [5] showed that if we have $\widehat{\boldsymbol{\Sigma}}\left(\mathbf{S}\right) = \mathbf{UL}^{-1}\boldsymbol{\Psi}^*(\mathbf{L})\mathbf{U}'$, where $\boldsymbol{\Psi}^*(\mathbf{L})$ is a diagonal matrix with ith diagonal elements equal to $\psi^*_i(\mathbf{L}), i = 1, \ldots, p$, then $\boldsymbol{\delta}_{EB}$ is minimax under the following conditions when the loss function is (2.2).

THEOREM 2.2. *Suppose that* $\psi^*_i(\mathbf{L})$, $i = 1, 2, \ldots, p$, *satisfy*:

(i) $0 < \psi^*_i(\mathbf{L}) < 2\,(m-p-1)$, $i = 1, 2, \ldots, p$;

(ii) *for every* $i = 1, 2, \ldots, p$, $\partial\psi^*_i(\mathbf{L})/\partial l_i \geqslant 0$;

(iii) $\psi^*_i(\mathbf{L})$ *are similarly ordered to* l_i, *that is*, $\big(\psi^*_i(\mathbf{L}) - \psi^*_t(\mathbf{L})\big)\left(l_i - l_t\right) \geqslant 0$ *for every* $1 \leqslant i, t \leqslant p$.

Then $\boldsymbol{\delta}_{EB} = \big(\mathbf{I}_p - \mathbf{VUL}^{-1}\boldsymbol{\Psi}^*(\mathbf{L})\,\mathbf{U}'\big)\,\mathbf{X}$ *improves on* $\mathbf{X}$ *for* $m > p+1$ *under the loss function* (2.2).

Proof. See Ghosh and Shieh [5]. ∎

By Theorem 2.1, we can conclude that

$$\boldsymbol{\delta}_{EB,\omega} = \big(\mathbf{I}_p - (1-\omega)\mathbf{VUL}^{-1}\boldsymbol{\Psi}^*(\mathbf{L})\mathbf{U}'\big)\mathbf{X}$$

is minimax for $m > p+1$ under the conditions of Theorem 2.2 when the loss function is (1.1). The examples (1) and (2) in Ghosh and Shieh [5] illustrated the above result. Of course, one may obtain estimators of the form (2.1) which dominate $\mathbf{X}$ when the conditions of Theorem 2.2 are not met. Some examples are given in Ghosh and Shieh [5].

In the rest of this section, we consider the problem of estimating $\boldsymbol{\Theta}$ under the loss function (1.1) for the case $\mathbf{Q} = \mathbf{V} = \mathbf{I}_p$. Let $\boldsymbol{\delta}_\omega = \mathbf{X} + (1-\omega)\mathbf{G}$, where $\mathbf{G} = \mathbf{G}(\mathbf{X})$ is a $p \times m$ matrix-valued function of $\mathbf{X}$. Also, let $\boldsymbol{\nabla}$ be a $p \times m$ matrix with the (i,j) element equal to the differential operator $\partial/\partial x_{i,j}$. We obtain a general condition for minimaxity.

THEOREM 2.3. *Suppose that* $\boldsymbol{\delta} = \mathbf{X} + \mathbf{G}$ *is minimax under the quadratic loss function* (2.2). *Then* $\boldsymbol{\delta}_\omega = \mathbf{X} + (1-\omega)\mathbf{G}$ *is minimax under the loss function* (1.1). *Also, an unbiased estimator of the risk of* $\boldsymbol{\delta}_\omega$ *under the loss function* (1.1) *is*

$$(1-\omega)[mp + (1-\omega)\{\operatorname{tr}(\mathbf{G}'\mathbf{G}) + 2\operatorname{tr}(\boldsymbol{\nabla}\mathbf{G}')\}].$$

P r o o f. We prove using a technique applied in Stein [13]. We have

$$\begin{aligned} R_1(\boldsymbol{\delta}_\omega;\boldsymbol{\Theta}) &= E_{\boldsymbol{\Theta}}[L_1(\boldsymbol{\delta}_\omega;\boldsymbol{\Theta})] \\ &= E_{\boldsymbol{\Theta}}\big[\omega\operatorname{tr}\big((\boldsymbol{\delta}_\omega - \mathbf{X})'(\boldsymbol{\delta}_\omega - \mathbf{X})\big) + (1-\omega)\operatorname{tr}\big((\boldsymbol{\delta}_\omega - \boldsymbol{\Theta})'(\boldsymbol{\delta}_\omega - \boldsymbol{\Theta})\big)\big] \\ &= R_1(\mathbf{X};\boldsymbol{\Theta}) + (1-\omega)^2 E_{\boldsymbol{\Theta}}\big[\operatorname{tr}(\mathbf{G}'\mathbf{G}) + 2\operatorname{tr}\big(\mathbf{G}'(\mathbf{X}-\boldsymbol{\Theta})\big)\big]. \end{aligned}$$

Using Stein's [13] identities, we can write

$$\begin{aligned} R_1(\boldsymbol{\delta}_\omega;\boldsymbol{\Theta}) - R_1(\mathbf{X};\boldsymbol{\Theta}) &= (1-\omega)^2 E_{\boldsymbol{\Theta}}[\operatorname{tr}(\mathbf{G}'\mathbf{G}) + 2\operatorname{tr}(\boldsymbol{\nabla}\mathbf{G}')] \\ &= (1-\omega)^2\big(R_3(\boldsymbol{\delta};\boldsymbol{\Theta}) - R_3(\mathbf{X};\boldsymbol{\Theta})\big). \end{aligned} \tag{2.5}$$

Hence, if $R_3(\boldsymbol{\delta};\boldsymbol{\Theta}) \leqslant R_3(\mathbf{X};\boldsymbol{\Theta})$, then $R_1(\boldsymbol{\delta}_\omega;\boldsymbol{\Theta}) \leqslant R_1(\mathbf{X};\boldsymbol{\Theta})$. Also, using (2.5), we have

$$R_1(\boldsymbol{\delta}_\omega;\boldsymbol{\Theta}) = (1-\omega)E_{\boldsymbol{\Theta}}[mp + (1-\omega)\{\operatorname{tr}(\mathbf{G}'\mathbf{G}) + 2\operatorname{tr}(\boldsymbol{\nabla}\mathbf{G}')\}].$$

Hence, $(1-\omega)[mp + (1-\omega)\{\operatorname{tr}(\mathbf{G}'\mathbf{G}) + 2\operatorname{tr}(\boldsymbol{\nabla}\mathbf{G}')\}]$ is an unbiased estimator of the risk of $\boldsymbol{\delta}_\omega$ under the loss function (2.5). The proof is complete. ∎

We can conclude from Theorem 2.3 that if $\{\operatorname{tr}(\mathbf{G}'\mathbf{G}) + 2\operatorname{tr}(\boldsymbol{\nabla}\mathbf{G}')\} \leqslant 0$ and a strict inequality holds with positive probability for some $\boldsymbol{\Theta}$, then $\boldsymbol{\delta}_\omega$ is minimax. Now, consider the class of shrinkage estimators

$$\boldsymbol{\delta}_{\omega,1} = \big(\mathbf{I}_p - (1-\omega)\mathbf{UF}^{-1}\boldsymbol{\Psi}(\mathbf{F})\mathbf{U}'\big)\mathbf{X}, \tag{2.6}$$

where $\boldsymbol{\Psi}(\mathbf{F}) = \operatorname{diag}(\psi_1,\ldots,\psi_p)$ is a diagonal matrix with ψ_i being functions of $\mathbf{F} = \mathbf{L}$. We obtain the following result by applying Theorem 2.3.

COROLLARY 2.1. *Suppose that* ψ_i *satisfy*:
(i) *for fixed* $f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_p$, $\partial\psi_i/\partial f_i \geqslant 0$, $i = 1, \ldots, p$;
(ii) $0 \leqslant \psi_p \leqslant \ldots \leqslant \psi_1 \leqslant 2\,(m-p-1)$.
Then $\boldsymbol{\delta}_{\omega,1}$ *in* (2.6) *is minimax under the loss function* (1.1).

Proof. With the notation $\mathbf{\Phi}(\mathbf{F}) = \mathbf{F}^{-1}\mathbf{\Psi}(\mathbf{F})$, Stein [13] showed that $\boldsymbol{\delta} = \big(\mathbf{I}_p - \mathbf{U}\mathbf{\Phi}(\mathbf{F})\mathbf{U}'\big)\mathbf{X}$ is minimax under the conditions (i) and (ii) when the loss function is (2.2). Using Theorem 2.3, we conclude that $\boldsymbol{\delta}_{\omega,1} = \mathbf{X} + (1-\omega)\mathbf{G}$, where $\mathbf{G} = -\mathbf{U}\mathbf{\Phi}(\mathbf{F})\mathbf{U}'\mathbf{X}$, is minimax under the loss function (1.1). ■

Tsukuma [15] showed that the proper Bayes estimators with respect to the following prior

$$\mathbf{\Theta} \sim N_{p\times m}\big(\mathbf{0}_{p\times m}, \mathbf{\Lambda}^{-1}\,(\mathbf{I}_p - \mathbf{\Lambda}) \otimes \mathbf{I}_m\big),$$

where $\mathbf{\Lambda}$ is a $p \times m$ random matrix with the density function

$$\pi\,(\mathbf{\Lambda}) \propto |\mathbf{\Lambda}|^{a/2-1}\,\mathbf{I}(\mathbf{0}_{p\times p} < \mathbf{\Lambda} < \mathbf{I}_p),$$

and $|\mathbf{\Lambda}|$ is the determinant of $\mathbf{\Lambda}$, are of the form $\boldsymbol{\delta} = \big(\mathbf{I}_p - \mathbf{U}\mathbf{F}^{-1}\mathbf{\Psi}\,(\mathbf{F})\,\mathbf{U}'\big)\,\mathbf{X}$. For certain a under the loss function (2.2), these estimators are minimax. For certain a under the loss function (1.1), the proper Bayes estimators are of the form $\boldsymbol{\delta}_{\omega,1} = \big(\mathbf{I}_p - (1-\omega)\mathbf{U}\mathbf{F}^{-1}\mathbf{\Psi}\,(\mathbf{F})\,\mathbf{U}'\big)\,\mathbf{X}$. So, Corollary 2.1 shows that $\boldsymbol{\delta}_{\omega,1}$ are minimax under the loss function (1.1).

3. MINIMAX ESTIMATION OF THE MEAN MATRIX UNDER THE LOSS FUNCTION (1.2)

Section 2 considered estimation of the mean matrix of a matrix-variate normal distribution. Here, we continue to discuss the problem of estimation under the loss function (1.2).

Lemma 3.1 shows that the maximum likelihood estimator is minimax under the loss function (1.2).

LEMMA 3.1. $\mathbf{X}$ *is minimax under the loss function* (1.2).

Proof. Note that the risk function of $\mathbf{X}$ is

$$R_2\,(\mathbf{X}, \mathbf{\Theta}) = E_{\mathbf{\Theta}}\,[L_2\,(\mathbf{X}; \mathbf{\Theta})] = k(1 - |\mathbf{V}|^{-m/2}\,|2\mathbf{\Gamma}^{-1} + \mathbf{V}^{-1}|^{-m/2}).$$

Now, suppose $\mathbf{\Theta}$ has a matrix-variate normal distribution with zero mean matrix and a covariance matrix $n\mathbf{C} \otimes \mathbf{I}_m$, i.e., $\mathbf{\Theta} \sim N_{p\times m}\,(\mathbf{0}, n\mathbf{C} \otimes \mathbf{I}_m)$. Then the posterior distribution is

$$\mathbf{\Theta}|\mathbf{X} \sim N_{p\times m}\bigg(\bigg(\mathbf{I}_p - \bigg(\mathbf{V}^{-1} + \frac{\mathbf{C}^{-1}}{n}\bigg)^{-1}\frac{\mathbf{C}^{-1}}{n}\bigg)\mathbf{X}, \bigg(\mathbf{V}^{-1} + \frac{\mathbf{C}^{-1}}{n}\bigg)^{-1} \otimes \mathbf{I}_m\bigg).$$

Thus, the Bayes estimator is

$$E_{\boldsymbol{\Theta}}\left[\boldsymbol{\Theta}|\,\mathbf{X}\right]=\left(\mathbf{I}_p-\left(\mathbf{V}^{-1}+\frac{\mathbf{C}^{-1}}{n}\right)^{-1}\frac{\mathbf{C}^{-1}}{n}\right)\mathbf{X}.$$

The corresponding Bayes risk is

$$r_n=k\left(1-\left|\mathbf{V}^{-1}+\frac{\mathbf{C}^{-1}}{n}\right|^{m/2}\left|2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1}+\frac{\mathbf{C}^{-1}}{n}\right|^{-m/2}\right).$$

Since determinants are polynomial functions of entries of a matrix, they are continuous functions, so

$$\lim_{n\to\infty}r_n=k(1-|\mathbf{V}|^{-m/2}\,|2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1}|^{-m/2})=\sup_{\boldsymbol{\Theta}}R_2(\mathbf{X};\boldsymbol{\Theta}).$$

Hence, by Theorem 1.12 of Lehmann and Casella [8], p. 316, and the results of Blyth [2], $\mathbf{X}$ is minimax under the loss function (1.2). ■

THEOREM 3.1. *If the estimators*

$$\boldsymbol{\delta}_2=\left(\mathbf{I}_p-(2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1})^{-1}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\right)\mathbf{X}$$

are minimax under the loss function (2.2) *with* $\mathbf{Q}=\boldsymbol{\Gamma}^{-1}$, *then they are also minimax under the loss function* (1.2), *where* $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ *depends on* $\mathbf{X}$ *only through* $\mathbf{S}$.

P r o o f. The difference of the risks of $\boldsymbol{\delta}_2$ and $\mathbf{X}$ is

$$R_2\left(\boldsymbol{\delta}_2;\boldsymbol{\Theta}\right)-R_2\left(\mathbf{X};\boldsymbol{\Theta}\right)=k[g\left(\mathbf{X},\boldsymbol{\Theta}\right)-g\left(\boldsymbol{\delta}_2,\boldsymbol{\Theta}\right)],$$

where

$$g\left(\boldsymbol{\delta},\boldsymbol{\Theta}\right)=E_{\boldsymbol{\Theta}}\left[e^{-\operatorname{tr}\left((\boldsymbol{\delta}-\boldsymbol{\Theta})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{\delta}-\boldsymbol{\Theta})\right)}\right].$$

Using $e^{-x}\geqslant 1-x$, we obtain

$$\begin{aligned}&g\left(\boldsymbol{\delta}_2,\boldsymbol{\Theta}\right)\geqslant g(\mathbf{X},\boldsymbol{\Theta})-|\mathbf{V}|^{-m/2}|2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1}|^{-m/2}\\&\times\left\{E_{\boldsymbol{\Theta}}\left[\operatorname{tr}\left(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})(2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1})^{-1}\boldsymbol{\Gamma}^{-1}(2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1})^{-1}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{X}\right)\right]\right.\\&\left.-\,2E_{\boldsymbol{\Theta}}\left[\operatorname{tr}\left(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})(2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1})^{-1}\boldsymbol{\Gamma}^{-1}(\mathbf{X}-\boldsymbol{\Theta})\right)\right]\right\},\end{aligned}$$

where $E_{\boldsymbol{\Theta}}$ is taken with respect to $N_{p\times m}\left(\boldsymbol{\Theta},(2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1})^{-1}\otimes\mathbf{I}_m\right)$. Using Theorem 2.1 in Ghosh and Shieh [5], we can write

$$\begin{aligned}(3.1)\quad &g\left(\boldsymbol{\delta}_2,\boldsymbol{\Theta}\right)-g(\mathbf{X},\boldsymbol{\Theta})\\&\geqslant-\,|\mathbf{V}|^{-m/2}\,|2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1}|^{-m/2}\left(R_3\left(\boldsymbol{\delta}_2;\boldsymbol{\Theta}\right)-R_3\left(\mathbf{X};\boldsymbol{\Theta}\right)\right).\end{aligned}$$

Hence, if $R_3\left(\boldsymbol{\delta}_2;\boldsymbol{\Theta}\right)\leqslant R_3\left(\mathbf{X};\boldsymbol{\Theta}\right)$, then $R_2\left(\boldsymbol{\delta}_2;\boldsymbol{\Theta}\right)\leqslant R_2\left(\mathbf{X};\boldsymbol{\Theta}\right)$. ■

By Ghosh and Shieh [5], we can write (3.1) as

$$g(\boldsymbol{\delta},\boldsymbol{\Theta})-g(\mathbf{X},\boldsymbol{\Theta})\geqslant -|\mathbf{V}|^{-m/2}\,|2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1}|^{-m/2}E_{\boldsymbol{\Theta}}[\mathrm{tr}(\mathbf{Q}_1^{1/2}\mathbf{U}_1\mathbf{Q}_1^{1/2})],$$

where $\mathbf{Q}_1=(2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1})^{-1}\boldsymbol{\Gamma}^{-1}(2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1})^{-1}$ and

$$\mathbf{U}_1=\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\mathbf{S}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})-4\mathbf{D}\big(\mathbf{S}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\big)-2(m-p-1)\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S}),$$

and $g(\boldsymbol{\delta},\boldsymbol{\Theta})-g(\mathbf{X},\boldsymbol{\Theta})\geqslant 0$ if $\mathbf{U}_1\leqslant 0$ has a positive probability for some $\boldsymbol{\Theta}$.

If $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})$ is chosen in such a way that $\boldsymbol{\delta}_2$ improves on $\mathbf{X}$ under the quadratic loss function (2.2), then $\boldsymbol{\delta}_2$ also improves on $\mathbf{X}$ under the loss function (1.2).

The following theorem illustrates the use of the above result.

THEOREM 3.2. *Suppose that* $\psi_i(\mathbf{L})$, $i=1,2,\ldots,p$, *satisfy*:

(i) $0<\psi_i(\mathbf{L})<2\,(m-p-1)$, $i=1,2,\ldots,p$;

(ii) *for every* $i=1,2,\ldots,p$, $\partial\psi_i(\mathbf{L})/\partial l_i\geqslant 0$;

(iii) $\psi_i(\mathbf{L})$ *are similarly ordered to* l_i, *that is*, $\big(\psi_i(\mathbf{L})-\psi_t(\mathbf{L})\big)\,(l_i-l_t)\geqslant 0$ *for every* $1\leqslant i,t\leqslant p$.

Then $\boldsymbol{\delta}_2^*(\mathbf{X})=\big(\mathbf{I}_p-(2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1})^{-1}\mathbf{U}\mathbf{L}^{-1}\boldsymbol{\Psi}\,(\mathbf{L})\,\mathbf{U}'\big)\mathbf{X}$ *improves on* $\mathbf{X}$ *for* $m>p+1$ *under the loss function* (1.2).

P r o o f. We use the proof of Theorem 2 in Ghosh and Shieh [5]. Note that if $\widehat{\boldsymbol{\Sigma}}^{-1}\,(\mathbf{S})=\mathbf{U}\boldsymbol{\Phi}\,(\mathbf{L})\,\mathbf{U}'$, where $\boldsymbol{\Phi}\,(\mathbf{L})=\mathbf{L}^{-1}\boldsymbol{\Psi}\,(\mathbf{L})$, we have

$$\mathbf{U}_1=\mathbf{U}\,[\boldsymbol{\Phi}\,(\mathbf{L})\,\mathbf{L}\boldsymbol{\Phi}\,(\mathbf{L})]\,\mathbf{U}'-4\mathbf{D}[\mathbf{U}\boldsymbol{\Phi}\,(\mathbf{L})\,\mathbf{L}\mathbf{U}']-2(m-p-1)\mathbf{U}\Phi\,(\mathbf{L})\,\mathbf{U}'.$$

By Haff [7], we have

$$\mathbf{D}[\mathbf{U}\boldsymbol{\Psi}\,(\mathbf{L})\,\mathbf{U}']=\mathbf{U}\boldsymbol{\Psi}^{(1)}\,(\mathbf{L})\,\mathbf{U}',$$

where $\boldsymbol{\Psi}^{(1)}\,(\mathbf{L})$ is a diagonal matrix with ith diagonal element equal to

$$\psi_i^{(1)}\,(\mathbf{L})=\frac{1}{2}\sum_{t\neq i}\frac{\psi_i\,(\mathbf{L})-\psi_t\,(\mathbf{L})}{l_i-l_t}+\frac{\partial\psi_i\,(\mathbf{L})}{\partial l_i}.$$

Therefore, $\mathbf{U}_1\leqslant\mathbf{0}$ is equivalent to $\mathbf{m}(\mathbf{L})\leqslant 0$, where $\mathbf{m}(\mathbf{L})$ is a diagonal matrix with ith diagonal element equal to

$$m_i(\mathbf{L})=\psi_i^2\,(\mathbf{L})\,l_i^{-1}-4\psi_i^{(1)}\,(\mathbf{L})-2(m-p-1)\psi_i\,(\mathbf{L})\,l_i^{-1}.$$

Now, by the assumptions (ii) and (iii), we have

$$\psi_i^{(1)}\,(\mathbf{L})\geqslant 0,\qquad i=1,\ldots,p.$$

Therefore,

$$m_i(\mathbf{L})\leqslant\psi_i^2\,(\mathbf{L})\,l_i^{-1}-2(m-p-1)\psi_i\,(\mathbf{L})\,l_i^{-1},$$

and by the assumption (i), we obtain $m_i(\mathbf{L})<0$. Hence, by Theorem 3.1, $\boldsymbol{\delta}_2^*(\mathbf{X})=\big(\mathbf{I}_p-(2\boldsymbol{\Gamma}^{-1}+\mathbf{V}^{-1})^{-1}\mathbf{U}\mathbf{L}^{-1}\boldsymbol{\Psi}\,(\mathbf{L})\,\mathbf{U}'\big)\mathbf{X}$ improves on $\mathbf{X}$ for $m>p+1$ under the loss function (1.2) and it is minimax. ∎

One can extend Theorem 3.1 as follows:

THEOREM 3.3. *The estimator*

$$\boldsymbol{\delta}_3 = \mathbf{X} + \mathbf{G}(\mathbf{X}),$$

where $\mathbf{G}(\mathbf{X})$ *is a* $p \times m$ *matrix-valued function of* $\mathbf{X}$, *is minimax under the loss function* (1.2) *if it is minimax under the loss function* (2.2).

Proof. The proof is similar to that of Theorem 3.1. ■

For the case $\mathbf{Q} = \mathbf{V} = \mathbf{I}_p$, Stein [13] showed that the estimator

$$\boldsymbol{\delta} = \left(\mathbf{I}_p - \mathbf{U}\mathbf{F}^{-1}\boldsymbol{\Psi}(\mathbf{F})\,\mathbf{U}'\right)\mathbf{X} \tag{3.2}$$

is minimax under the loss function (2.2). Using Theorem 3.3, we can see that the estimator (3.2) is minimax under the loss function (1.2).

4. A SIMULATION STUDY

Here, we compare the performances of the estimators given by Theorems 2.1 and 3.1. The performance was based on a simulation study and was assessed with respect to bias and mean squared error.

We simulated ten thousand samples each of size $m = 5, 6, \ldots, 100$ from a matrix-variate normal distribution with zero means and covariance matrix $\mathbf{V} \otimes \mathbf{I}_m$. We computed the biases and mean squared errors for the estimators given by Theorems 2.1 and 3.1 for each m. Figure 1 plots the biases versus m. Figure 2 plots the mean squared errors versus m.

We can see that there is not much to choose between the estimators with respect to bias. The estimator given by Theorem 2.1 has consistently smaller mean squared error for every m. However, both estimators behave as expected: their biases approach zero with increasing m and their mean squared errors decrease to zero with increasing m. Hence, either of the two estimators can be chosen if bias is important. The estimator given by Theorem 2.1 should be preferred if mean squared error is important.

For the simulations in Figures 1 and 2, we took $\mathbf{Q} = \mathbf{I}_p$, $\mathbf{V} = \mathbf{I}_p$, $\boldsymbol{\Gamma} = \mathbf{I}_p$ and $\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S}) = (m - p - 1)\mathbf{S}^{-1}$. The simulation results were similar for a wide range of other choices for $\mathbf{Q}$, $\mathbf{V}$ and $\boldsymbol{\Gamma}$. In particular, there was not much to choose between the estimators with respect to bias, the estimator given by Theorem 2.1 always had consistently smaller mean squared error for every m, the biases of both estimators always approached zero with increasing m and the mean squared errors of both estimators always decreased to zero with increasing m.

The fact that the estimator in Theorem 3.1 does not perform well does not mean that there is no need to study the estimation of mean under the loss function (1.2).

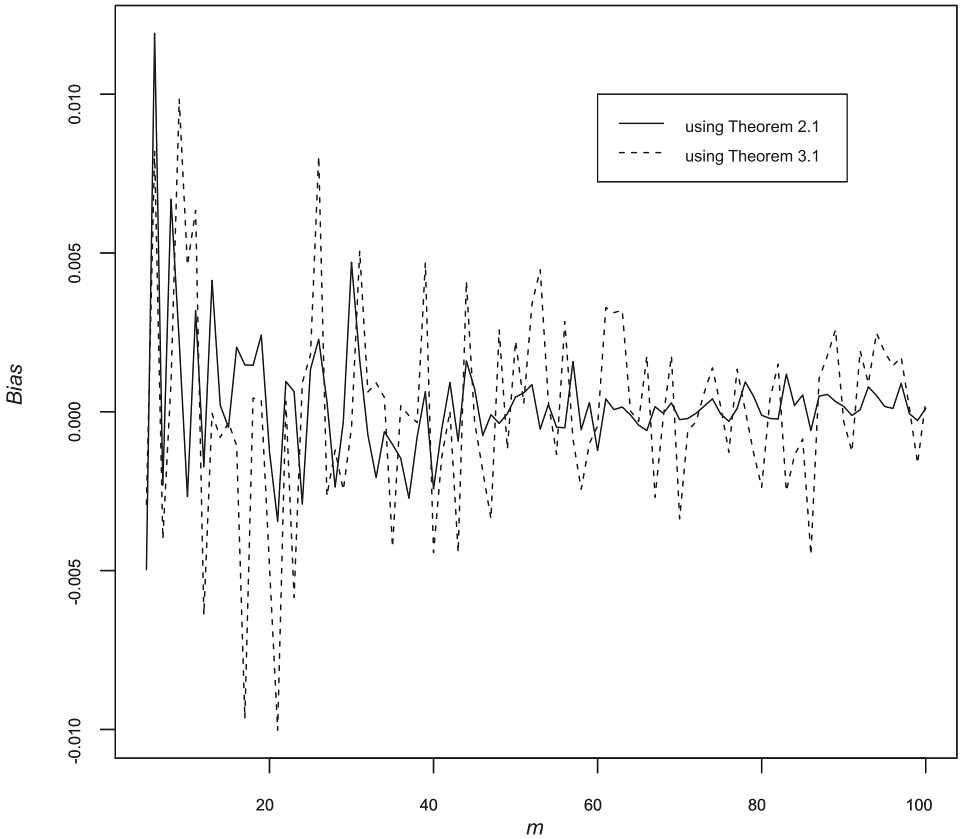

FIGURE 1. Biases of the estimators given by Theorems 2.1 and 3.1 versus m

We could not have known that the estimator in Theorem 3.1 does not perform well without first deriving the estimators of the mean under the two loss functions. Besides, there may be other parameters or parameters of other matrix-variate distributions for which the loss function (1.2) may provide better estimators.

The derivations given in Sections 2 and 3 are illustrations on how estimators of a parameter of a particular matrix-variate distribution can be derived under the two loss functions. The derivations can help the reader to derive estimators (under the two loss functions) for other parameters or parameters of other matrix-variate distributions.

Finally, we would like to say that both Theorems 2.1 and 3.1 give estimators better than the maximum likelihood estimator of the mean. The fact that Theorem 3.1 gives a better estimator can be proved as follows: by Ghosh and Shieh [5], it follows that if $\widehat{\mathbf{\Sigma}}^{-1}(\mathbf{S}) = (m-p-1)b\mathbf{S}^{-1} + \mathbf{S}/\operatorname{tr}(\mathbf{S}^2)$, then

$$\psi_i(\mathbf{L}) = m - p - 1 + \frac{bl_i^2}{\sum_{j=1}^p l_j^2}.$$

It is easy to verify conditions (i)–(iii) of Theorem 3.2 for this particular choice of

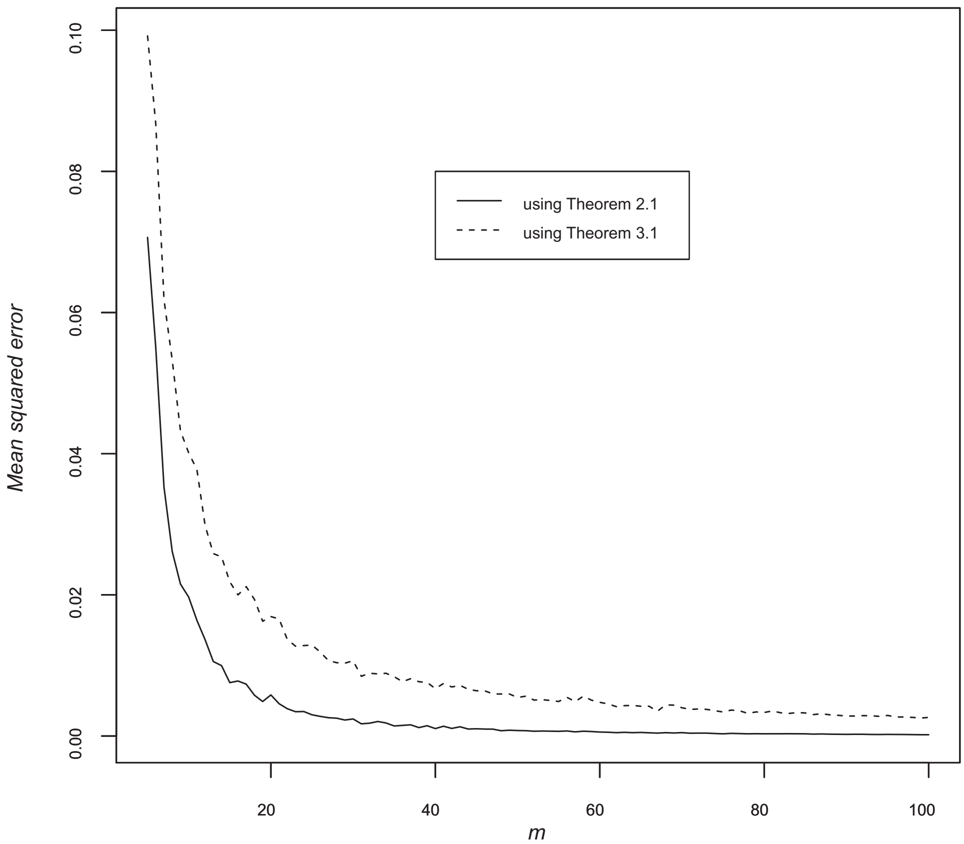

FIGURE 2. Mean squared errors of the estimators given by Theorems 2.1 and 3.1 versus m

$\boldsymbol{\Psi}$ for $b \in [0, m-p-1)$. Thus, $\big(\mathbf{I}_p - (2\boldsymbol{\Gamma}^{-1} + \mathbf{V}^{-1})^{-1}\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{S})\big)\mathbf{X}$ dominates $\mathbf{X}$ for the loss function (1.2), so it is minimax. Hence, the proof.

Acknowledgments. The authors would like to thank the two referees for careful reading and comments which greatly improved the paper.

REFERENCES

[1] A. Asgharazadeh and N. S. Farsipour, *Estimation of the multivariate normal mean under the extended balanced loss function*, Internat. J. Statist. Systems 3 (2008), pp. 131–136.

[2] C. R. Blyth, *On minimax statistical decision procedures and their admissibility*, Ann. Math. Statist. 22 (1951), pp. 22–42.

[3] D. K. Dey, M. Ghosh, and W. E. Strawderman, *On estimation with balanced loss functions*, Statist. Probab. Lett. 45 (1999), pp. 97–101.

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

[5] M. Ghosh and G. Shieh, *Empirical Bayes minimax estimators of matrix normal means*, J. Multivariate Anal. 38 (1991), pp. 306–318.

[6] R. L. Haff, *Further identities for the Wishart distribution with applications in regression*, Canad. J. Statist. 9 (1981), pp. 215–244.
[7] R. L. Haff, *Solutions of the Euler–Lagrange equations for certain multivariate normal estimation problems*, unpublished (1984).
[8] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, second edition, Springer, New York 1998.
[9] B. P. K. Leung and F. A. Spiring, *Some properties of the family of inverted probability loss functions*, Quality Technology and Quantitative Management 1 (2004), pp. 125–147.
[10] K. D. Majeske and T. W. Lauer, *The bank loan approval decision from multiple perspectives*, Expert Systems with Applications 40 (2013), pp. 1591–1598.
[11] J. N. Pan, *A new loss function-based method for evaluating manufacturing and environmental risks*, International Journal of Quality and Reliability Management 24 (2007), pp. 861–887.
[12] F. A. Spiring, *The reflected normal loss function*, Canad. J. Statist. 21 (1993), pp. 321–330.
[13] C. Stein, *Estimation of the mean of a multivariate normal distribution*, in: *Proceedings of the Prague Symposium on Asymptotic Statistics, September 3–6, 1973*, J. Hájek (Ed.), Prague 1974, pp. 345–381.
[14] M. Towhidi and J. Behboodian, *Estimation of the multivariate normal mean under the extended reflected normal loss functions*, Bull. Iranian Math. Soc. 28 (2002), pp. 57–65.
[15] H. Tsukuma, *Admissibility and minimaxity of Bayes estimators for a normal mean matrix*, J. Multivariate Anal. 99 (2008), pp. 2251–2264.
[16] A. Zellner, *Bayesian and non-Bayesian estimation using balanced loss functions*, in: *Statistical Decision Theory and Related Topics*, S. S. Gupta and J. O. Berger (Eds.), Springer, New York 1994, pp. 371–390.
[17] Z. Zhang, *On estimation of matrix of normal mean*, J. Multivariate Anal. 18 (1986), pp. 70–82.
[18] Z. Zhang, *Selecting a minimax estimator doing well, at a point*, J. Multivariate Anal. 19 (1986), pp. 14–23.

S. Zinodiny
Department of Mathematics
and Computer Science
Amirkabir University of Technology
Tehran, Iran
E-mail: shokofeh_zinodiny@yahoo.com

S. Rezaei
Department of Mathematics
and Computer Science
Amirkabir University of Technology
Tehran, Iran
E-mail: srezaei@aut.ac.ir

S. Nadarajah
School of Mathematics
University of Manchester
Manchester M13 9PL, United Kingdom
E-mail: mbbsssn2@manchester.ac.uk

Received on 5.5.2015;
revised version on 20.9.2015