

A class of Bivariate SURE estimators in heteroscedastic hierarchical normal models

S.K. Ghoreishi

*Department of Statistics, Faculty of Sciences, University of Qom, Qom, Iran
atty_ghoreishi@yahoo.com*

"Tgegkxgf "33"lcpwct {"4239
""Ceeegr vgf "7"lvpq"4239

In this paper, we first propose a class of bivariate shrinkage estimators based on Steins unbiased estimate of risk (SURE). Then, we study the effect of correlation coefficients on their performance. Moreover, under some mild assumptions on the model correlations, we set up the optimal asymptotic properties of our estimates when the number of vector means to be estimated grows. Furthermore, we carry out a simulation study to compare how various bivariate competing shrinkage estimators perform and analyze a real data set.

Keywords: Asymptotic univariate shrinkage estimators; Heteroscedasticity; Hierarchical models; Multivariate shrinkage estimator; Stein's unbiased risk estimators.

2000 Mathematics Subject Classification: 62F15, 62F30

1. Introduction

One of the most applicable class of models is the class of hierarchical models. It is well-known that the shrinkage estimates play an important role in hierarchical models. In this line, the James-Stein estimators, which combine the partial information of various sources, dominates some other estimators like the ordinary least squares estimators, Stein (1956) and James and Stein (1961). However, the good risk properties of shrinkage estimators make them appealing in many applicable disciplines. Undoubtedly, the basic works of James and Stein on shrinkage estimators became a foundation for the other progressing studies on hierarchical normal models. Stein (1962) described a hierarchical, empirical Bayes interpretation for the shrinkage estimators. Further studies have been done by Efron and Morris (1973), who tried to interpret these kinds of empirical Bayes estimators by several other competing parametric empirical Bayes estimators. Moreover, the homoscedastic (equal subpopulation variances) empirical Bayes interpretation of estimators as well as heteroscedastic (unequal subpopulation variances) ones has motivated various treatments of this problem in hierarchical normal models. For homoscedastic case, we can refer to Baranchik (1970), Strawderman (1971), Brown (1971, 1975) and Berger (1976) among the others, while the heteroscedastic situation have been addressed by a few authors. For more details see Hudson (1974), Xie et al. (2012), Ghoreishi and Meshkani (2014, 2015).

Following considerable work on univariate shrinkage estimators, the multivariate type of James-Stein shrinkage estimators have also been considered for random vectors of dimension $r \geq 3$, which either shrink towards the origin vector or another given fixed vector, Lehmann and Casella (1998). James and Stein (1961) showed that the multiple homoscedastic shrinkage estimator of the population mean dominates the corresponding maximum likelihood estimator when $r \geq 3$. It means that the multiple James and Stein shrinkage estimator always achieves lower MSE than the corresponding maximum likelihood estimator. Equivalently, the James and Stein multiple estimator makes the maximum likelihood estimator inadmissible when $r \geq 3$.

From interpretation point of view, the James and Stein estimator, as an empirical Bayes, is based on the assumption that the population mean is a random vector with an unknown diagonal covariance matrix $\sigma^2 I$ and this matrix is estimated from the data itself, Brown (2008), Brown and Greenshtein (2009) and Berger and Strawderman (2009). The main point on this approach is that it is achievable when the random vector of means has a dimension greater than or equal 3 and unfortunately it is not applicable for 2-dimensional random vectors.

The study of James and Stein (1961) on empirical Bayes shrinkage estimators have extended to the case of a general covariance matrix, i.e., where measurements are statistically dependent and may have different variances, see Strawderman (1971) and Bock (1975). They illustrated that the theoretical results on multivariate shrinkage estimator hold when $r \geq 3$ and in addition the covariance matrix has the general form $\sigma^2 \mathbf{D}$, which is known up to constant σ^2 .

To our knowledge, all results on multivariate homoscedastic James-Stein shrinkage estimators hold for large dimensions $r \geq 3$.

The outlines of the paper are:

- 1) We first introduce an alternative class of bivariate shrinkage estimators.
- 2) We study the correlation effect on their performance.
- 3) Since the correlation structure is the biggest challenge in bivariate setting, and this phenomenon can seriously threaten the truth of theoretical results, under some mild assumptions on the model correlations, we show that our proposed estimators have optimal asymptotic properties.

The paper is organized as follows. In Section 2, we review some basic concepts and define necessary notations. The main results of this work, including the risk properties of our bivariate conditional heteroscedastic SURE estimators, are presented in Section 3. We carry out a simple simulation study to evaluate the performance of our proposed bivariate empirical Bayes shrinkage estimators in Section 4 and illustrate the theoretical results for a real data set in Section 5. Necessary technical proofs are given in the appendix.

2. Preliminaries

Let $\mathbf{Y}_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix}$ be the vectors, with 2×1 dimension, of normal random variables taken independently at points $i = 1, \dots, N$. That is

$$\mathbf{Y}_i \sim N_2(\boldsymbol{\theta}_i, \mathbf{A}_i). \quad (2.1)$$

Moreover, The classical conjugate hierarchical model puts a prior of bivariate normal distribution on θ_i

$$\theta_i \sim N_2(\mu, \Lambda).$$

Combination of two assumed model leads us to the following heteroscedastic hierarchical bivariate normal model

$$\begin{aligned} \mathbf{Y}_i &\sim N_2(\theta_i, \mathbf{A}_i), \\ \theta_i &\sim N_2(\mu, \Lambda), \quad i = 1, \dots, N, \end{aligned} \tag{2.2}$$

where $\Lambda = \text{diag}\{\lambda_i\}$ is an unknown 2×2 diagonal matrix with positive entries, μ is a 2×1 unknown vector and \mathbf{A}_i s are (potentially) distinct covariance matrices either completely known or will be known by some plug-in robust estimators.

Application of Bayes theorem for model (2.2) gives us the bivariate posterior density

$$\theta_i \sim N_2((\mathbf{A}_i^{-1} + \Lambda^{-1})^{-1}(\mathbf{A}_i^{-1}\mathbf{Y}_i + \Lambda^{-1}\mu), (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-1}).$$

According to this distribution, we consider the posterior means as an alternative class of bivariate shrinkage estimators of θ_i . That is,

$$\hat{\theta}_i^{\hat{\Lambda}, \hat{\mu}} = (\mathbf{A}_i^{-1} + \hat{\Lambda}^{-1})^{-1}\mathbf{A}_i^{-1}\mathbf{Y}_i + (\mathbf{A}_i^{-1} + \hat{\Lambda}^{-1})^{-1}\hat{\Lambda}^{-1}\hat{\mu}, \tag{2.3}$$

As one can see these estimators tend to shrink the sample vectors to the mean vector μ via some coefficients, having important roles to balance observations effect in estimation.

It is straightforward to check that the following relationship holds between the coefficients of estimators (2.3),

$$(\mathbf{A}_i^{-1} + \hat{\Lambda}^{-1})^{-1}\mathbf{A}_i^{-1} = I_{2 \times 2} - (\mathbf{A}_i^{-1} + \hat{\Lambda}^{-1})^{-1}\hat{\Lambda}^{-1}, \tag{2.4}$$

for identity matrix $I_{2 \times 2}$. Moreover, we have

$$\begin{aligned} (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-1}\mathbf{A}_i^{-1} &= (I_{2 \times 2} + \mathbf{A}_i\Lambda^{-1})^{-1} \\ &= (I_{2 \times 2} + \begin{pmatrix} a_{1i}^2 & \rho_i a_{1i} a_{2i} \\ \rho_i a_{1i} a_{2i} & a_{2i}^2 \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{pmatrix})^{-1} \\ &= \begin{pmatrix} U_{11i}^\Lambda & -U_{12i}^\Lambda \\ -U_{21i}^\Lambda & U_{22i}^\Lambda \end{pmatrix} = \mathbf{U}_i^\Lambda, \end{aligned}$$

and

$$(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-1}\Lambda^{-1} = \begin{pmatrix} 1 - U_{11i}^\Lambda & U_{12i}^\Lambda \\ U_{21i}^\Lambda & 1 - U_{22i}^\Lambda \end{pmatrix},$$

where $a_{1i} > 0, a_{2i} > 0$ and

$$\begin{aligned} U_{11i}^\Lambda &= \frac{1 + \frac{a_{2i}^2}{\lambda_2}}{(1 + \frac{a_{1i}^2}{\lambda_1})(1 + \frac{a_{2i}^2}{\lambda_2}) - \rho_i^2 \frac{a_{1i}^2 a_{2i}^2}{\lambda_1 \lambda_2}}, & U_{12i}^\Lambda &= \frac{\frac{\rho_i a_{1i} a_{2i}}{\lambda_1}}{(1 + \frac{a_{1i}^2}{\lambda_1})(1 + \frac{a_{2i}^2}{\lambda_2}) - \rho_i^2 \frac{a_{1i}^2 a_{2i}^2}{\lambda_1 \lambda_2}} \\ U_{21i}^\Lambda &= \frac{\frac{\rho_i a_{1i} a_{2i}}{\lambda_2}}{(1 + \frac{a_{1i}^2}{\lambda_1})(1 + \frac{a_{2i}^2}{\lambda_2}) - \rho_i^2 \frac{a_{1i}^2 a_{2i}^2}{\lambda_1 \lambda_2}}, & U_{22i}^\Lambda &= \frac{1 + \frac{a_{1i}^2}{\lambda_1}}{(1 + \frac{a_{1i}^2}{\lambda_1})(1 + \frac{a_{2i}^2}{\lambda_2}) - \rho_i^2 \frac{a_{1i}^2 a_{2i}^2}{\lambda_1 \lambda_2}}. \end{aligned}$$

It is worth pointing out the following remark concerning our shrinkage estimators.

Remark 2.1. Condition $|\rho_i| \rightarrow 0$ is equivalent to assume separate shrinkage estimates for marginal variables. If $\lambda_1, \lambda_2 \rightarrow \infty$, then our shrinkage estimates severely shrank to the observations. Moreover, if $|\rho_i| > \frac{\min\{a_{1i}, a_{2i}\}}{\max\{a_{1i}, a_{2i}\}}$, under some extreme values for the other parameters, our shrinkage estimates behave in an uncontrolled manner.

Since the heteroscedastic model, though more appropriate for practical applications, is less well studied in higher dimensions, and it is unclear what types of bivariate shrinkage estimators are superior in terms of the risk, in this work, we propose SURE approach under some mild conditions on correlation coefficients. For our purpose, we adopt the quadratic loss function

$$l_N(\theta, \hat{\theta}^{\Lambda, \mu}) = \frac{1}{N} \sum_{i=1}^N \{\theta_i - \hat{\theta}_i^{\Lambda, \mu}\}' \{\theta_i - \hat{\theta}_i^{\Lambda, \mu}\}.$$

Let us denote the corresponding risk function and its unbiased estimator by $R(\theta, \hat{\theta}^{\Lambda, \mu})$ and $SURE(\Lambda, \mu)$, respectively.

For Bivariate shrinkage estimators (2.3), we have

$$l_N(\theta, \hat{\theta}^{\Lambda, \mu}) = \frac{1}{N} \sum_{i=1}^N \{(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-1} \mathbf{A}_i^{-1} (\mathbf{Y}_i - \theta_i) + (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-1} \Lambda^{-1} (\mu - \theta_i)\}' \\ \times \{(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-1} \mathbf{A}_i^{-1} (\mathbf{Y}_i - \theta_i) - (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-1} \Lambda^{-1} (\theta_i - \mu)\}.$$

Therefore, the corresponding risk function is obtained as follows

$$R(\theta, \hat{\theta}^{\Lambda, \mu}) = E(l_N(\theta, \hat{\theta}^{\Lambda, \mu})) \\ = \frac{1}{N} \sum_{i=1}^N \{tr[\mathbf{A}_i^{-1} (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2}] + (\theta_i - \mu)' \Lambda^{-1} (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2} \Lambda^{-1} (\theta_i - \mu)\}.$$

It is Straightforward to see that the Stein's unbiased risk estimator of the $R(\theta, \hat{\theta}^{\Lambda, \mu})$ is given by

$$SURE(\Lambda, \mu) = \frac{1}{N} \sum_{i=1}^N \{tr[\mathbf{A}_i^{-1} (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2} \mathbf{A}_i^{-1} - \Lambda^{-1} (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2} \Lambda^{-1}] \mathbf{A}_i + \\ (\mathbf{Y}_i - \mu)' \Lambda^{-1} (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2} \Lambda^{-1} (\mathbf{Y}_i - \mu)\},$$

for more details see the appendix. In terms of $U_{11i}, U_{12i}, U_{21i}$, and U_{22i} , $SURE(\Lambda, \mu)$ can be rewritten as

$$SURE(\Lambda, \mu) = \frac{1}{N} \sum_{i=1}^N \{[(1 - U_{11i})^2 + U_{21i}^2] (Y_{1i} - \mu_1)^2 + [U_{12i}^2 + (1 - U_{22i})^2] (Y_{2i} - \mu_2)^2 + \\ 2[(1 - U_{11i})U_{12i} + U_{21i}(1 - U_{22i})] (Y_{1i} - \mu_1)(Y_{2i} - \mu_2) + \\ (1 - 2U_{11i})a_{1i}^2 + 2\rho_i a_{1i} a_{2i} [U_{12i} + U_{21i}] + a_{2i}^2 (1 - 2U_{22i})\}.$$

To estimate μ_1, μ_2, λ_1 , and λ_2 , one can minimize $SURE(\Lambda, \mu)$ with respect to these quantities. Setting $\frac{\partial SURE(\Lambda, \mu)}{\partial \mu_1} = 0$ and $\frac{\partial SURE(\Lambda, \mu)}{\partial \mu_2} = 0$, give the solutions, satisfying

$$\frac{1}{N} \sum_{i=1}^N [(1 - U_{11i})^2 + U_{21i}^2] \mu_1 + \frac{1}{N} \sum_{i=1}^N [(1 - U_{11i})U_{12i} + U_{21i}(1 - U_{22i})] \mu_2 = \quad (2.5)$$

$$\frac{1}{N} \sum_{i=1}^N \{[(1 - U_{11i})^2 + U_{21i}^2]Y_{1i} + [(1 - U_{11i})U_{12i} + U_{21i}(1 - U_{22i})]Y_{2i}\}$$

and

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N [(1 - U_{11i})U_{12i} + U_{21i}(1 - U_{22i})]\mu_1 + \frac{1}{N} \sum_{i=1}^N [U_{12i}^2 + (1 - U_{22i})^2]\mu_2 = \quad (2.6) \\ \frac{1}{N} \sum_{i=1}^N \{[(1 - U_{11i})U_{12i} + U_{21i}(1 - U_{22i})]Y_{1i} + [U_{12i}^2 + (1 - U_{22i})^2]Y_{2i}\}. \end{aligned}$$

In order to elicit the solutions, we assume $(\lambda_1, \lambda_2) \in [0, b]^2$, for relatively large real value $b > 0$. Then, we partition this cell into several tiny cells and apply (2.5) and (2.6) for (λ_1, λ_2) at each grid of these tiny cells and simultaneously compute μ_1, μ_2 and their corresponding SURE. The values of μ_1, μ_2, λ_1 , and λ_2 that minimize the SURE function are considered the SURE estimates. So, the SURE approach leads us to the Stein' unbiased risk estimators of Λ and μ which can obtain by minimizing $SURE(\Lambda, \mu)$ with respect to the Λ and μ , when the solutions exist. That is,

$$(\hat{\Lambda}_{SURE}, \hat{\mu}_{SURE}) = \min \arg_{\Lambda, \mu} SURE(\Lambda, \mu).$$

In this case, the corresponding SURE estimators are of the form

$$\hat{\theta}_i^{\hat{\Lambda}_{SURE}, \hat{\mu}_{SURE}} = (\mathbf{A}_i^{-1} + \hat{\Lambda}_{SURE}^{-1})^{-1} \mathbf{A}_i^{-1} \mathbf{Y}_i + (\mathbf{A}_i^{-1} + \hat{\Lambda}_{SURE}^{-1})^{-1} \hat{\Lambda}_{SURE}^{-1} \hat{\mu}_{SURE}. \quad (2.7)$$

In the next section, we discuss how our SURE estimators $\hat{\theta}_i^{\hat{\Lambda}_{SURE}}$ and μ_{SURE} perform in comparison with other competitive estimators resulting from other approaches including Empirical Bayes Method of Moment Estimators (EBMME) and Empirical Bayes Maximum Likelihood Estimators (EBMLE).

3. Asymptotic properties of the bivariate SURE estimators

In this section, we show that the proposed bivariate SURE estimators (2.7) have optimal asymptotic properties in comparison with their competitor bivariate estimators. In order to check how well $SURE(\Lambda, \mu)$ function approximates the loss function $l_N(\theta_i, \hat{\theta}_i^{\Lambda, \mu})$, we have

$$\begin{aligned} SURE(\Lambda, \mu) - l_N(\theta, \hat{\theta}^{\Lambda, \mu}) &= \frac{1}{N} \sum_{i=1}^N \{Y_{1i}^2 + Y_{2i}^2 - a_{1i}^2 - a_{2i}^2 - \theta_{1i}^2 - \theta_{2i}^2\} \\ &\quad - \frac{2}{N} \sum_{i=1}^N \{U_{11i}^\Lambda Y_{1i}^2 + U_{22i}^\Lambda Y_{2i}^2 - (U_{12i}^\Lambda + U_{21i}^\Lambda) Y_{1i} Y_{2i} \\ &\quad - U_{11i}^\Lambda Y_{1i} \theta_{1i} - U_{22i}^\Lambda Y_{2i} \theta_{2i} + U_{12i}^\Lambda Y_{2i} \theta_{1i} + U_{21i}^\Lambda Y_{1i} \theta_{2i} \\ &\quad - U_{11i}^\Lambda a_{1i}^2 - U_{22i}^\Lambda a_{2i}^2 + (U_{12i}^\Lambda + U_{21i}^\Lambda) \rho_i a_{1i} a_{2i}\} \\ &\quad - \frac{2}{N} \sum_{i=1}^N \{(\mu_1(1 - U_{11i}^\Lambda) + \mu_2 U_{21i}^\Lambda)(Y_{1i} - \theta_{1i}) \\ &\quad + (\mu_2(1 - U_{22i}^\Lambda) + \mu_1 U_{12i}^\Lambda)(Y_{2i} - \theta_{2i})\}. \end{aligned}$$

By rearranging the terms of the above equation, one has

$$\begin{aligned}
 SURE(\Lambda, \mu) - l_N(\theta, \hat{\theta}^{\Lambda, \mu}) &= \underbrace{\frac{1}{N} \sum_{i=1}^N \{Y_{1i}^2 - a_{1i}^2 - \theta_{1i}^2\}}_{I_1} - \frac{2}{N} \sum_{i=1}^N U_{11i}^\Lambda \{Y_{1i}^2 - Y_{1i}\theta_{1i} - a_{1i}^2\} \\
 &\quad - \underbrace{\frac{2}{N} \sum_{i=1}^N \mu_1 (1 - U_{11i}^\Lambda)(Y_{1i} - \theta_{1i})}_{I_2} - \underbrace{\frac{2}{N} \sum_{i=1}^N \mu_2 U_{21i}^\Lambda (Y_{1i} - \theta_{1i})}_{I_3} \\
 &\quad + \underbrace{\left[\frac{1}{N} \sum_{i=1}^N \{Y_{2i}^2 - a_{2i}^2 - \theta_{2i}^2\} - \frac{2}{N} \sum_{i=1}^N U_{22i}^\Lambda \{Y_{2i}^2 - Y_{2i}\theta_{2i} - a_{2i}^2\} \right]}_{II_1} \\
 &\quad - \underbrace{\frac{2}{N} \sum_{i=1}^N \mu_2 (1 - U_{22i}^\Lambda)(Y_{2i} - \theta_{2i})}_{II_2} - \underbrace{\frac{2}{N} \sum_{i=1}^N \mu_1 U_{12i}^\Lambda (Y_{2i} - \theta_{2i})}_{II_3} \\
 &\quad + \underbrace{\frac{2}{N} \sum_{i=1}^N U_{12i}^\Lambda \{Y_{1i}Y_{2i} - Y_{2i}\theta_{1i} - \rho_i a_{1i}a_{2i}\}}_{III_1} \\
 &\quad + \underbrace{\frac{2}{N} \sum_{i=1}^N U_{21i}^\Lambda \{Y_{1i}Y_{2i} - Y_{1i}\theta_{2i} - \rho_i a_{1i}a_{2i}\}}_{III_2}.
 \end{aligned} \tag{3.1}$$

To investigate the risk properties of our bivariate SURE estimators, we need to use Lemma (2.1) in Li(1986). In our setting, it is applicable when the absolute values of U_i^Λ entries are less than or equal to 1. Hence, the following lemma gives some mild conditions on correlation coefficients to guarantee the framework of Lemma (2.1) in Li(1986). Since the proof is straightforward, it is omitted.

Lemma 3.1. For each t , the elements of U_i^Λ , satisfy in

- i) $0 \leq U_{11i}^\Lambda \leq 1$,
- ii) $0 \leq U_{22i}^\Lambda \leq 1$,
- iii) If $|\rho_i| \leq \frac{\min\{a_{1i}, a_{2i}\}}{\max\{a_{1i}, a_{2i}\}}$, then $|U_{12i}^\Lambda| \leq 1$, $|U_{21i}^\Lambda| \leq 1$, and $|U_{12i}^\Lambda + U_{21i}^\Lambda| \leq 1$.

In Lemma (3.1)[iii], the upper bound for the absolute value of the correlation coefficients can be removed when $a_{1i} = a_{2i}$. Under this condition, both variables have the same variances and so the quantities U_{12i}^Λ , U_{21i}^Λ , and ρ_i have the same sign.

To proceed, we assume a *homogeneous property* for the correlation coefficients. It means either $\rho_i \geq 0$, for all $i = 1, 2, \dots, N$; or $\rho_i \leq 0$, for all $i = 1, 2, \dots, N$; which is a natural assumption for any relation between two assumed variables over sub-populations. For example, in a multinomial distribution, which can be approximated by a multivariate normal distribution, every two cell frequencies always has a negative correlation. With this assumption, the asymptotic property of our bivariate SURE function, $SURE(\Lambda, \mu)$, is given by the following theorem. The necessary conditions (C1-C3) are stated in the appendix.

Theorem 3.1. For model (2.2), under the conditions C1-C3 and homogeneous property of the correlation coefficients, If $|\rho_i| \leq \frac{\min\{a_{1i}, a_{2i}\}}{\max\{a_{1i}, a_{2i}\}}$, then

$$\sup_{\Lambda, \mu} |SURE(\Lambda, \mu) - l_N(\theta, \hat{\theta}_i^{\Lambda, \mu})| \rightarrow 0$$

in L^1 and in probability, as $N \rightarrow \infty$.

Another class of bivariate shrinkage estimators, which have appealing asymptotic properties, is the shrinkage estimators that are shrunk toward the data grand mean. That is,

$$\hat{\theta}_i^{\Lambda, \bar{Y}} = (\mathbf{A}_i^{-1} + \hat{\Lambda}^{-1})^{-1} \mathbf{A}_i^{-1} \mathbf{Y}_i + (\mathbf{A}_i^{-1} + \hat{\Lambda}^{-1})^{-1} \hat{\Lambda}^{-1} \bar{\mathbf{Y}}. \tag{3.2}$$

Using quadratic loss function for these bivariate estimators, it is straightforward to see that

$$\begin{aligned} SURE(\Lambda) - l_N(\theta_i, \hat{\theta}_i^{\Lambda, \bar{Y}}) &= \underbrace{\frac{1}{N} \sum_{i=1}^N \{Y_{1i}^2 - a_{1i}^2 - \theta_{1i}^2\} - \frac{2}{N} \sum_{i=1}^N U_{11i}^{\Lambda} \{Y_{1i}^2 - Y_{1i}\theta_{1i} - a_{1i}^2\}}_{I_1} \\ &\quad - \underbrace{\frac{2}{N} \sum_{i=1}^N (1 - U_{11i}^{\Lambda}) \{\bar{Y}_1(Y_{1i} - \theta_{1i}) - \frac{a_{1i}^2}{N}\}}_{I_2'} \\ &\quad - \underbrace{\frac{2}{N} \sum_{i=1}^N U_{21i}^{\Lambda} \{\bar{Y}_2(Y_{1i} - \theta_{1i}) - \frac{\rho_i a_{1i} a_{2i}}{N}\}}_{I_3'} \\ &\quad + \underbrace{\frac{1}{N} \sum_{i=1}^N \{Y_{2i}^2 - a_{2i}^2 - \theta_{2i}^2\} - \frac{2}{N} \sum_{i=1}^N U_{22i}^{\Lambda} \{Y_{2i}^2 - Y_{2i}\theta_{2i} - a_{2i}^2\}}_{II_1} \\ &\quad - \underbrace{\frac{2}{N} \sum_{i=1}^N (1 - U_{22i}^{\Lambda}) \{\bar{Y}_2(Y_{2i} - \theta_{2i}) - \frac{a_{2i}^2}{N}\}}_{II_2'} \\ &\quad - \underbrace{\frac{2}{N} \sum_{i=1}^N U_{12i}^{\Lambda} \{\bar{Y}_1(Y_{2i} - \theta_{2i}) - \frac{\rho_i a_{1i} a_{2i}}{N}\}}_{II_3'} \\ &\quad + \underbrace{\frac{2}{N} \sum_{i=1}^N U_{12i}^{\Lambda} \{Y_{1i}Y_{2i} - Y_{2i}\theta_{1i} - \rho_i a_{1i} a_{2i}\}}_{III_1} \\ &\quad + \underbrace{\frac{2}{N} \sum_{i=1}^N U_{21i}^{\Lambda} \{Y_{1i}Y_{2i} - Y_{1i}\theta_{2i} - \rho_i a_{1i} a_{2i}\}}_{III_2}. \end{aligned} \tag{3.3}$$

Therefore, the following theorem holds for (3.3).

Theorem 3.2. For model (2.2), under the conditions C1-C3 and homogenous property for the correlation coefficients, If $|\rho_i| \leq \frac{\min\{a_{1i}, a_{2i}\}}{\max\{a_{1i}, a_{2i}\}}$, then

$$\sup_{\Lambda} |SURE(\Lambda) - l_N(\theta, \hat{\theta}_i^{\Lambda, \bar{Y}})| \rightarrow 0$$

in L^1 and in probability, as $N \rightarrow \infty$.

Remark 3.1. Here, it is important to note that if $|\rho_i| \rightarrow 0$, the upper bound of (A.1), in the appendix, tends to become sharper. This means that under uncorrelated situation the asymptotic convergence rate increases in Theorems (3.1) and (3.2).

4. Simulation study

In this section, we carry out a simple simulation to study how our bivariate shrinkage estimates perform in comparison with EBMME (Empirical Bayes Method of Moments Estimator) and EBMLE (Empirical Bayes Maximum Likelihood Estimator) bivariate estimators defined below.

- Empirical Bayes method of moment estimators:

$$\hat{\theta}_i^{\hat{\Lambda}_{EBMME}, \hat{\mu}_{EBMME}} = (\mathbf{A}_i^{-1} + \hat{\Lambda}_{EBMME}^{-1})^{-1} \mathbf{A}_i^{-1} \mathbf{Y}_i + (\mathbf{A}_i^{-1} + \hat{\Lambda}_{EBMME}^{-1})^{-1} \hat{\Lambda}_{EBMME}^{-1} \hat{\mu}_{EBMME},$$

where $\hat{\Lambda}_{EBMME}$ and $\hat{\mu}_{EBMME}$ satisfy in

$$\begin{aligned} \sum_{i=1}^N (\mathbf{A}_i + \Lambda)^{-\frac{1}{2}} (\mathbf{Y}_i - \mu) &= 0, \\ \sum_{i=1}^N \{(\mathbf{Y}_i - \mu)' (\mathbf{A}_i + \Lambda)^{-1} (\mathbf{Y}_i - \mu) - 2\} &= 0, \end{aligned}$$

and the elements of $\hat{\Lambda}_{EBMME}$ are non-negative.

- Empirical Bayes maximum likelihood estimators:

$$\hat{\theta}_i^{\hat{\Lambda}_{EBMLE}, \hat{\mu}_{EBMLE}} = (\mathbf{A}_i^{-1} + \hat{\Lambda}_{EBMLE}^{-1})^{-1} \mathbf{A}_i^{-1} \mathbf{Y}_i + (\mathbf{A}_i^{-1} + \hat{\Lambda}_{EBMLE}^{-1})^{-1} \hat{\Lambda}_{EBMLE}^{-1} \hat{\mu}_{EBMLE},$$

where the elements of $\hat{\Lambda}_{EBMLE}$ and $\hat{\mu}_{EBMLE}$ are obtained by maximizing the log-likelihood function

$$\begin{aligned} l = cte - \frac{1}{2} \sum_{i=1}^N \ln\{(a_{1i}^2 + \lambda_1)(a_{2i}^2 + \lambda_2) - \rho_i^2 a_{1i}^2 a_{2i}^2\} \\ - \sum_{i=1}^N \frac{1}{2\{(a_{1i}^2 + \lambda_1)(a_{2i}^2 + \lambda_2) - \rho_i^2 a_{1i}^2 a_{2i}^2\}} \{(a_{2i}^2 + \lambda_2)(Y_{1i} - \mu_1)^2 \\ + (a_{1i}^2 + \lambda_1)(Y_{2i} - \mu_2)^2 - 2\rho_i a_{1i} a_{2i} (Y_{1i} - \mu_1)(Y_{2i} - \mu_2)\}. \end{aligned}$$

For our simulation study, we first derive samples for

$$\mathbf{A}_i = \begin{pmatrix} a_{1i}^2 & \rho_i a_{1i} a_{2i} \\ \rho_i a_{1i} a_{2i} & a_{2i}^2 \end{pmatrix}$$

with marginal distributions $a_{1i}^2 \sim Unif(1, 2)$ and $a_{2i}^2 \sim Unif(0.1, 1)$, and compute the corresponding correlation coefficients under three following scenarios.

- Scenario I: we assume a large correlation coefficient for each $i = 1, 2, \dots, N$. That is $\rho_i = 0.95 \frac{\min(a_{1i}, a_{2i})}{\max(a_{1i}, a_{2i})}$.
- Scenario II: we assume a moderate correlation coefficient, $\rho_i = 0.25 \frac{\min(a_{1i}, a_{2i})}{\max(a_{1i}, a_{2i})}$.
- Scenario III: A weak correlation coefficient is considered in this scenario, $\rho_i = 0.05 \frac{\min(a_{1i}, a_{2i})}{\max(a_{1i}, a_{2i})}$.

We also assume that θ_i are independently distributed according to

$$\theta_i \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}\right).$$

To measure the predictive power of different bivariate shrinkage estimators, we compute the limiting risk $\lim_{N \rightarrow \infty} R(\theta, \hat{\theta})$ for different shrinkage estimators under three scenarios. This process is repeated a large number of times $T = 10,000$ to obtain an accurate estimate of the average risk for each estimator under three scenarios. Moreover, the sample size N is chosen to be 50 and 200. Table 1 shows the result for various approaches.

Table 1. The risk of various bivariate shrinkage estimates

Sample Size	Method	Scenario I	Scenario II	Scenario III
$N = 50$	Oracle	0.3342	0.1021	0.0971
	SURE	0.3343	0.1044	0.0978
	EBMME	0.4558	0.3322	0.2320
	EBMLE	0.4865	0.3413	0.2403
$N = 200$	Oracle	0.1947	0.0540	0.0097
	SURE	0.1989	0.0553	0.0098
	EBMME	0.2271	0.0653	0.0187
	EBMLE	0.2430	0.0670	0.0188

Table 1 clearly confirms that although all bivariate shrinkage estimates (SURE, EBMLE and EBMME) have good performance in comparison with the corresponding Oracle risk estimates, our bivariate SURE estimates do better in terms of the risk value under Scenarios (I)-(III). That is, as our theoretical result indicates, the performance of the SURE estimators approaches that of the oracle risk estimators.

5. Application to a real dataset

Educational performance is one of the principle measures for evaluation of an educational center, like a university. It is usually considered through two extreme groups of the students who have either excellent performance (with grade point averages between [17,20]) or poor performance (with grade point averages less than 12) in comparison with the third (moderate) group (with grade point averages between [12,17)). In the following, we use E , M , and P notations which stand for groups of students with Excellent, Moderate, and Poor performances. For a given department indexed by i , let E_i , M_i , and P_i denote the number of students at each group and let n_i denote the number of total students in the corresponding department during the given semester. Whenever it is necessary to consider multiple semesters of educations, such as the two semesters in a given educational year,

we will insert other subscripts, and use symbols such as E_{ji} , M_{ji} , and P_{ji} , to denote the number of students with excellent, moderate, and poor performances and total number n_{ij} for department i ($= 1, \dots, N$) within semester j ($= 1, 2$).

For data involving N departments over two semesters, we focus on two extreme groups E and P , which are known as educational outcome of each department, Tinto (2010), and assume that

$$(E_{ji}, B_{ji}) \sim \text{multinomial}(n_{ji}, p_{ji}^E, p_{ji}^P); i = 1, \dots, N; j = 1, 2,$$

where unobserved parameters p_{ji}^E and p_{ji}^P are the corresponding proportions of groups E and P with the assumptions of $p_{1i}^E = p_{2i}^E$ and $p_{1i}^P = p_{2i}^P$. Furthermore, the corresponding sample proportions are denoted by $\hat{p}_{ji}^E = \frac{E_{ji}}{n_{ji}}$ and $\hat{p}_{ji}^P = \frac{P_{ji}}{n_{ji}}$. It is obvious that the sample proportions have approximately multivariate normal with vector mean (p_{ji}^E, p_{ji}^P) , and a covariance matrix that depends on the unknown values p_{ji}^E and p_{ji}^P .

For our purpose, we are interested in dealing with nearly bivariate normal variables such that a) having variances that depend only on the observed value of n and b) asymptotic performance of bivariate shrinkage estimators are robust with respect to plug-in estimates for the correlation coefficient. From Bartlett (1936, 1947), the standard variance-stabilizing transformations, $\arcsin \sqrt{\frac{E}{n}}$ and $\arcsin \sqrt{\frac{P}{n}}$ may do this goal fairly well for univariate case. However, in our bivariate setting, we prefer to use the transformations

$$Y_1 = \arcsin \sqrt{\frac{E + 1/4}{n + 1/2}},$$

and

$$Y_2 = \arcsin \sqrt{\frac{P + 1/4}{n + 1/2}}.$$

For more details on these transformations, see Brown (2008). It is well-known that these transformations have marginal variance-stabilizing properties. That is:

$$E(Y_1) := \arcsin \sqrt{p^E}, \text{ var}(Y_1) = \frac{1}{4n}$$

$$E(Y_2) := \arcsin \sqrt{p^P}, \text{ var}(Y_2) = \frac{1}{4n}$$

By applying Delta-method for the function $\arcsin \sqrt{x}$, it is straightforward to see that the estimated correlation between Y_1 and Y_2 is given by

$$\hat{\rho}_{Y_1, Y_2} = -\sqrt{\frac{(E + 1/4)(P + 1/4)}{(n - E + 1/4)(n - P + 1/4)}} + O\left(\frac{1}{n^2}\right)$$

We would like to note that Lemma (3.1) holds for this correlation and the class of correlations given by (5.1). Furthermore, it is not difficult to numerically verify that the asymptotic performance of our

proposed bivariate shrinkage estimators are robust with respect to a general class of the correlation between Y_1 and Y_2

$$\hat{\rho}_{Y_1, Y_2}^c = -\sqrt{\frac{(E+c)(P+c)}{(n-E+c)(n-P+c)}}, \tag{5.1}$$

for a positive constant $c \ll n$. For more details on marginal transformation of type (5.1), see Brown (2008). If in equation (5.1) we put $c = 0$, the estimated correlation will be equal to the maximum likelihood estimate of correlation between the sample proportions. In summary, we approximately have

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \arcsin \sqrt{p^E} \\ \arcsin \sqrt{p^B} \end{pmatrix}, \frac{1}{4n} \begin{pmatrix} 1 & \hat{\rho}_{Y_1, Y_2} \\ \hat{\rho}_{Y_1, Y_2} & 1 \end{pmatrix}\right)$$

Let’s look at Educational data, Table 2, for 24 departments at University of Qom over the educational year 2014. Our primary focus in this research is to evaluate the students’ performance in the second semester, at each department, when the parameters are estimated based on the first-semester data. Since we use the information of an already-completed educational year, we can then validate our bivariate estimation performance by comparing the estimated values with the real values for the second semester. To evaluate an estimator $\hat{\theta}$ based on the transformation \mathbf{Y}_{2j} , we measure the total bivariate sum of squared prediction errors(TSE) defined as

$$TSE(\hat{\theta}) = \sum_{i=1}^N (\mathbf{Y}_{2i} - \hat{\theta}_i)' (\mathbf{Y}_{2i} - \hat{\theta}_i) - \sum_{i=1}^N \frac{1}{2n_{2i}}.$$

Table 3 summarizes the numerical results of our bivariate shrinkage estimators in comparison with the other competing bivariate shrinkage estimators. The corresponding TSE values are reported. Acceptable performance of our bivariate SURE estimators, among the other estimators, are shown by Table 3. To numerically investigate the robust performance of our bivariate shrinkage estimators, with respect to little changes in correlation coefficient, we present a plot of TSE versus c in (5.1). As Figure 1 illustrates, TSE values of two approaches EBMME and EBMLE always dominate the corresponding values of SURE even for fairly large value of c .

Acknowledgements

I am grateful to Editor, Associate Editor and three referees for their constructive and insightful comments that helped greatly improve the quality of the paper.

Table 2. Number of students according to their educational performance and department of study at each semester over educational year 2014 in University of Qom

Department	Semester 1				Semester 2			
	E	M	P	Total	E	M	P	Total
Statistics	5	16	1	22	4	13	4	21
Law	28	21	2	51	25	22	4	51
Mathematics	5	18	8	31	4	13	12	29
English literature	5	29	9	43	9	18	14	41
Arabic literature	23	15	3	41	14	16	9	41
Persian literature	24	29	3	56	11	30	10	51
Biology	4	7	2	13	2	9	1	12
Chemistry	14	28	7	49	8	31	9	48
Information	16	16	1	33	23	8	1	32
Economics	9	21	6	36	3	22	11	36
Education	20	21	3	44	13	25	6	44
Quranic Sciences	23	3	1	27	17	8	2	27
Computer Sciences	6	23	6	35	3	16	14	32
Jurisprudence	32	15	0	47	38	7	2	47
Philosophy	23	8	4	35	23	5	3	31
Physics	6	27	6	39	3	20	15	38
Business Management	20	18	6	44	20	16	8	44
Industrial Management	27	4	0	31	19	12	1	32
Architectural Engineering	13	5	1	19	1	16	2	19
Electrical Engineering	23	26	1	50	12	33	5	50
Industrial Engineering	4	13	2	19	6	12	1	19
Civil Engineering	25	28	8	61	13	40	8	61
Computer Engineering	18	24	2	44	15	27	1	43
Mechanical Engineering	16	11	3	30	8	17	5	30

Table 3. TSE for various c

c	0.25	1	2	5	10
SURE	0.101	0.079	0.064	0.035	0.021
EBMME	0.520	0.459	0.388	0.249	0.163
EBMLE	0.557	0.507	0.426	0.281	0.192

Appendix

We first prove $E(SURE(\Lambda, \mu)) = R(\theta, \hat{\theta}^{\Lambda, \mu})$. It is easy to see

$$E(SURE(\Lambda, \mu)) = \frac{1}{N} \sum_{i=1}^N \{tr[\mathbf{A}_i^{-1}(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2} \mathbf{A}_i^{-1} - \Lambda^{-1}(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2} \Lambda^{-1}] \mathbf{A}_i + E[(\mathbf{Y}_i - \mu)' \Lambda^{-1} (\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2} \Lambda^{-1} (\mathbf{Y}_i - \mu)]\}$$

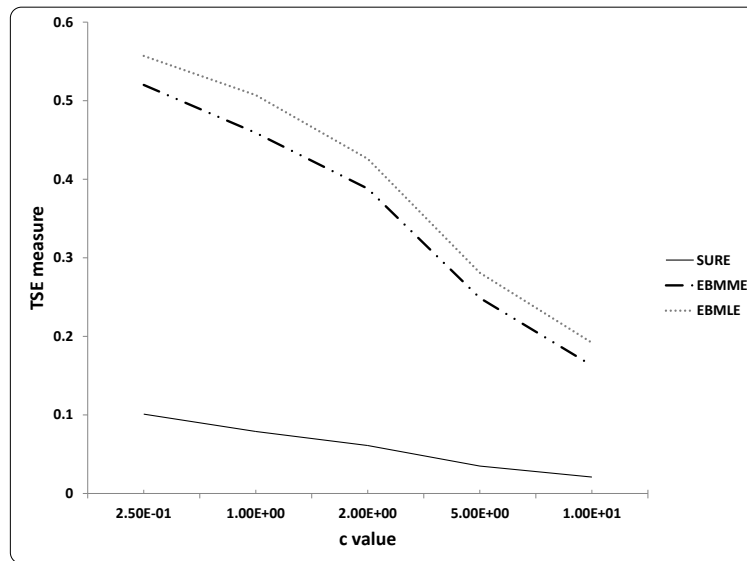


Fig. 1. TSE against c values

$$\begin{aligned}
 &= \frac{1}{N} \sum_{i=1}^N \{tr[\mathbf{A}_i^{-1}(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2}\mathbf{A}_i^{-1} - \Lambda^{-1}(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2})\Lambda^{-1}]\mathbf{A}_i \\
 &\quad + [\boldsymbol{\theta}_i - \boldsymbol{\mu}]' \Lambda^{-1}(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2}\Lambda^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu}) \\
 &\quad + tr[\Lambda^{-1}(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2})\Lambda^{-1}\mathbf{A}_i]\} \\
 &= \frac{1}{N} \sum_{i=1}^N \{tr(\mathbf{A}_i^{-1}(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2}) + (\boldsymbol{\theta}_i - \boldsymbol{\mu})' \Lambda^{-1}(\mathbf{A}_i^{-1} + \Lambda^{-1})^{-2}\Lambda^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu})\} \\
 &= R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{\Lambda, \boldsymbol{\mu}}). \square
 \end{aligned}$$

For establishing the asymptotic results, the following conditions are required,

C1) For $j = 1, 2$ and some $\delta > 0$:

$$\begin{cases} \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_{ji}^4 < \infty, \\ \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_{ji}^2 \theta_{ji}^2 < \infty, \\ \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N a_{1i}^2 \theta_{2i}^2 < \infty, \\ \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N a_{2i}^2 \theta_{1i}^2 < \infty. \end{cases}$$

C2)

$$\begin{cases} \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \theta_{ji}^{2+\delta} < \infty. \\ \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \theta_{1i} < \infty, \\ \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \theta_{2i} < \infty, \end{cases}$$

C3) Let $|\boldsymbol{\mu}_1| \leq \max_t |Y_{1t}|$, and $|\boldsymbol{\mu}_2| \leq \max_t |Y_{2t}|$. Since the sensible shrinkage estimators would attempt to shrink toward locations that lie inside the range of the data, these assumptions are included for technical reasons to facilitate the proof. These conditions imply that

$E(\max_{1 \leq i \leq N} Y_{ji}^2) = O(N^{\frac{2}{2+\delta^*}})$, for $j = 1, 2$, where $\delta^* = \min(1, \delta)$, which is necessary to prove our theorems. For more details see Xie et al. (2012).

Proof of Theorem (3.1)

From (3.1), we have

$$\begin{aligned} \sup_{\Lambda, \mu} |SURE(\Lambda, \mu) - l_N(\theta, \hat{\theta}_i^{\Lambda, \mu})| &\leq \sup_{\Lambda, \mu} |I_1| + \sup_{\Lambda, \mu} |I_2| + \sup_{\Lambda, \mu} |I_3| \\ &\quad + \sup_{\Lambda, \mu} |II_1| + \sup_{\Lambda, \mu} |II_2| + \sup_{\Lambda, \mu} |II_3| \\ &\quad + \sup_{\Lambda, \mu} |III_1| + \sup_{\Lambda, \mu} |III_2|. \end{aligned}$$

The L^2 convergence of the $\sup_{\Lambda, \mu} |I_1|$ and $\sup_{\Lambda, \mu} |I_2|$, and L^1 convergence of $\sup_{\Lambda, \mu} |I_3|$, $\sup_{\Lambda, \mu} |II_2|$, and $\sup_{\Lambda, \mu} |III_3|$ immediately obtain from Theorems (3.1) and (5.1) in Xie et al. (2012) for univariate case. The proof of two last expressions, which consider interaction between two random variables, is given below. Without loss of generality, we assume wherever it is needed, the summands are rearranged according to decreasing order of their coefficients U_{12i}^Λ . Therefore, we have

$$\begin{aligned} \sup_{\Lambda, \mu} |III_1| &= \sup_{\Lambda} \left| \frac{2}{N} \sum_{i=1}^N U_{12i}^\Lambda \{Y_{1i}Y_{2i} - Y_{2i}\theta_{1i} - \rho_i a_{1i}a_{2i}\} \right| \\ &= \sup_{\lambda_2 > 0} \sup_{\lambda_1 > 0} \left| \frac{2}{N} \sum_{i=1}^N U_{12i}^\Lambda \{Y_{1i}Y_{2i} - Y_{2i}\theta_{1i} - \rho_i a_{1i}a_{2i}\} \right|. \end{aligned}$$

Under necessary conditions for correlation coefficients in Lemma (3.1)[iii], and using Lemma (2.1) in Li (1986), the last term is equal to

$$\begin{aligned} &\sup_{\{1 \geq U_{121}^\Lambda \geq U_{122}^\Lambda, \dots, \geq U_{12N}^\Lambda \geq 0\}} \left| \frac{2}{N} \sum_{i=1}^N U_{12i}^\Lambda \{Y_{1i}Y_{2i} - Y_{2i}\theta_{1i} - \rho_i a_{1i}a_{2i}\} \right| \\ &\leq \max_{1 \leq t \leq N} \left| \frac{2}{N} \sum_{i=1}^t \{Y_{1i}Y_{2i} - Y_{2i}\theta_{1i} - \rho_i a_{1i}a_{2i}\} \right|. \end{aligned}$$

Let $M_t = \sum_{i=1}^t \{Y_{1i}Y_{2i} - Y_{2i}\theta_{1i} - \rho_i a_{1i}a_{2i}\}$. It is easy to see that $M_t; t = 1, 2, \dots$ form a martingale process. Therefore the L^2 maximum inequality implies

$$\begin{aligned} E(\max_{1 \leq t \leq N} M_t^2) &\leq 4E(M_N^2) = 4E\left[\sum_{i=1}^N \{(Y_{1i} - \theta_{1i})Y_{2i} - \rho_i a_{1i}a_{2i}\}^2\right] \\ &= 4E\left[\sum_{i=1}^N \{(Y_{1i} - \theta_{1i})^2 Y_{2i}^2 + \rho_i^2 a_{1i}^2 a_{2i}^2 - 2\rho_i a_{1i}a_{2i}(Y_{1i} - \theta_{1i})Y_{2i}\}\right] \\ &= 4 \sum_{i=1}^N [(1 + \rho_i^2) a_{1i}^2 a_{2i}^2 + a_{1i}^2 \theta_{2i}^2]. \end{aligned} \tag{A.1}$$

Therefore, conditions (C1) and (C2) imply that $E(\max_{1 \leq t \leq N} (\frac{2}{N} M_t)^2) \rightarrow 0$, which is equivalent to

$$\sup_{\Lambda, \mu} |III_1| = \sup_{\Lambda} \left| \frac{2}{N} \sum_{i=1}^N U_{12i}^\Lambda \{Y_{1i}Y_{2i} - Y_{2i}\theta_{1i} - \rho_i a_{1i}a_{2i}\} \right| \rightarrow 0 \text{ in } L^2 \text{ as } N \rightarrow \infty.$$

The same argument shows

$$\sup_{\Lambda, \mu} |III_2| = \sup_{\Lambda} \left| \frac{2}{N} \sum_{i=1}^N U_{12i}^{\Lambda} \{Y_{1i}Y_{2i} - Y_{1i}\theta_{2i} - \rho_i a_{1i}a_{2i}\} \right| \rightarrow 0 \text{ in } L^2 \text{ as } N \rightarrow \infty. \square$$

Proof of Theorem (3.2)

Form (3.3) it is obvious that

$$\begin{aligned} \sup_{\Lambda} |SURE(\Lambda) - l_N(\theta_i, \hat{\theta}_i^{\Lambda, \bar{Y}})| &\leq \sup_{\Lambda} |I_1| + \sup_{\Lambda} |I'_2| + \sup_{\Lambda} |I'_3| \\ &\quad + \sup_{\Lambda} II_1 + \sup_{\Lambda} |II'_2| + \sup_{\Lambda} |II'_3| \\ &\quad + \sup_{\Lambda} |III_1| + \sup_{\Lambda} |III_2|. \end{aligned}$$

For univariate L^2 convergence of the $\sup_{\Lambda} |I_1|$ and $\sup_{\Lambda} |II_2|$, L^1 convergence of $\sup_{\Lambda} |I'_2|$ and $\sup_{\Lambda} |II'_2|$ see Xie et al. (2012). The proof of L^2 convergence of $\sup_{\Lambda} |III_1|$ and $\sup_{\Lambda} |III_2|$ is discussed in Theorem (3.1). The proof of convergence of interaction terms $\sup_{\Lambda} |I'_3|$ and $\sup_{\Lambda} |II'_3|$ comes below. To prove the convergence of $\sup_{\Lambda} |I'_3|$, we assume the summands are rearranged according to decreasing order of their coefficients. Therefore, we have

$$\begin{aligned} \sup_{\Lambda} |I'_3| &= \sup_{\Lambda} \left| \frac{2}{N} \sum_{i=1}^N U_{21i}^{\Lambda} \left\{ \bar{Y}_2(Y_{1i} - \theta_{1i}) - \frac{\rho_i a_{1i} a_{2i}}{N} \right\} \right| \\ &\leq \sup_{\Lambda} \left| \frac{2}{N} \sum_{i=1}^N U_{21i}^{\Lambda} \bar{Y}_2(Y_{1i} - \theta_{1i}) \right| + \sup_{\Lambda} \left| \frac{2}{N} \sum_{i=1}^N \frac{\rho_i a_{1i} a_{2i}}{N} \right| \\ &\leq \sup_{\Lambda} \left| \frac{2}{N} \sum_{i=1}^N U_{21i}^{\Lambda} (Y_{1i} - \theta_{1i}) \right| |\bar{Y}_2| + \sup_{\Lambda} \left| \frac{2}{N^2} \sum_{i=1}^N a_{1i} a_{2i} \right|. \end{aligned}$$

Following the technique in the proof of Theorem (3.1), it is obvious to see that

$$\sup_{\Lambda} \left| \frac{2}{N} \sum_{i=1}^N U_{21i}^{\Lambda} (Y_{1i} - \theta_{1i}) \right| \rightarrow 0 \text{ in } L^2.$$

Moreover, $E(\bar{Y}_2^2) = \frac{1}{N^2} \sum_{i=1}^N a_{2i}^2 + (\frac{1}{N} \sum_{i=1}^N \theta_{2i})^2$, which is bounded by conditions C1-C3. So, by Cauchy-Schwartz inequality we have

$$\sup_{\Lambda} \left| \frac{2}{N} \sum_{i=1}^N U_{21i}^{\Lambda} (Y_{1i} - \theta_{1i}) \right| |\bar{Y}_2| \rightarrow 0 \text{ in } L^1.$$

Therefore, we have

$$\sup_{\Lambda} |I'_3| \rightarrow 0 \text{ in } L^1.$$

The same results hold for $\sup_{\Lambda} |II'_3|$ and hence this completes the proof. \square

References

- [1] A.J. Baranchik, A Family of Minimax Estimators of the Mean of a Multivariate Normal Distribution, *Ann. of Math. Statist.* **41** (1970) 642–645.
- [2] J. Berger, Admissible Minimax Estimation of a Multivariate Normal Mean With Arbitrary Quadratic Loss, *Ann. of Statist.* **4** (1976) 223–226.
- [3] J. Berger and W.E. Strawderman, Choice of Hierarchical Priors: Admissibility in Estimation of Normal Means, *Ann. of Statist.* **24** (1996) 931–951.
- [4] M.E. Bock, Minimax Estimators of the Mean of a Multivariate Normal Distribution, *Ann. of Statist.* **3** (1975) 209–218.
- [5] L.D. Brown, Admissible Estimators, Recurrent Diffusions, and Insoluble Boundary Value Problems, *Ann. of Math. Statist.* **42** (1971) 855–903.
- [6] L.D. Brown, Estimation With Incompletely Specified Loss Functions (the Case With Several Location Parameters), *J. Amer. Statist. Assoc.* **70** (1975) 417–427.
- [7] L.D. Brown, In-Season Prediction of Batting Average: A Field Test of Empirical Bayes and Bayes Methodologies, *Ann. of Appl. Statist.* **2** (2008) 113–152.
- [8] L.D. Brown and E. Greenshtein, Nonparametric Empirical Bayes and Compound Decision Approaches to Estimation of a High-Dimensional Vector of Means, *Ann. of Statist.* **37** (2009) 1685–1704.
- [9] B. Efron and C. Morris, Stein’s Estimation Rule and Its Competitors: An Empirical Bayes Approach, *J. Amer. Statist. Assoc.* **68** (1973) 117–130.
- [10] S.K. Ghoreishi, Bayesian analysis of hierarchical heteroscedastic linear models using Dirichlet-Laplace priors, *J. of Statistical Theory and Applications* **16** (2017) 53–64.
- [11] S.K. Ghoreishi and M.R. Meshkani, On SURE estimators in hierarchical models assuming heteroscedasticity for both levels of a two-level normal hierarchical model, *J. of Multivariate Analysis* **132** (2014) 129–137.
- [12] S.K. Ghoreishi and A. Mostafavinia, Shrinkage estimators for multi-level heteroscedastic hierarchical normal linear models, *J. of Statistical Theory and Applications* **14** (2015) 204–213.
- [13] H.M. Hudson, Empirical Bayes Estimation, Technical Report No. 58. Department of Statistics, Stanford University 1974.
- [14] W. James and C.M. Stein, Estimation With Quadratic Loss, *Proc. 4th Berkeley Symposium on Probability and Statistics I* (1961) 367–379.
- [15] E.L. Lehmann and G. Casella, *Theory of Point Estimation* (2nd Ed.), Springer, New York 1998.
- [16] K.C. Li, Asymptotic optimality of C_L and Generalized Cross Validation in Ridge Regression with Application to Spline Smoothing, *Ann. of Statist.* **14** (1986) 1101–1112.
- [17] C.M. Stein, Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** (1956) 197–206.
- [18] C.M. Stein, Confidence Sets for the Mean of a Multivariate Normal Distribution (with discussion), *J. Roy. Statist. Soc. Ser. B.* **24** (1962) 265–296.
- [19] W.E. Strawderman, Proper Bayes Minimax Estimators of the Multivariate Normal Mean, *Ann. Math. Statist.* **42** (1971) 385–388.
- [20] X. Xie, S.C. Kou and L.D. Brown, SURE estimators for a Heteroscedastic Hierarchical Model, *J. Amer. Statist. Assoc.* **107** (2012) 1465–1479.