97-F-20

Double Shrinkage Estimation of Common Coefficients in Two Regression Equations with Heteroscedasticity

Tatsuya Kubokawa, Faculty of Economics
University of Tokyo

June 1997

Double Shrinkage Estimation of Common Coefficients in Two Regression Equations with Heteroscedasticity

Tatsuya KUBOKAWA

Department of Economics, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

The problem of estimating the common regression coefficients is addressed in this paper for two regression equations with possibly different error variances. The feasible generalized least squares (FGLS) estimators have been believed to be admissible within the class of unbiased estimators. It is, nevertheless, established that the FGLS estimators are inadmissible in the light of minimizing the covariance matrices if the dimension of the common regression coefficients is greater than or equal to three. Double shrinkage unbiased estimators are proposed as possible candidates of improved procedures.

AMS 1991 subject classifications. Primary 62C15; secondary 62F11, 62H12, 62J05.

Key words and phrases. Common mean problem, feasible (two-stage) generalized least squares estimators, inadmissibility, unbiased estimation, heteroscedastic linear regression model.

1. Introduction

Consider the problem of estimating common regression coefficients β of two linear regression equations

$$y_i = X_i\beta + e_i, \quad i = 1, 2$$

where y_i is an $n_i \times 1$ vector of observations, X_i is an $n_i \times p$ known matrix of rank p and e_i is an $n_i \times 1$ random vector having normal distribution $\mathcal{N}_{n_i}(0, \sigma_i^2 I_{n_i})$ for $n_i \times n_i$ identity matrix I_{n_i}. Let β be a $p \times 1$ vector of unknown common regression coefficients and let σ_1^2 and σ_2^2 be unknown dispersion parameters possibly different. In this model, the minimal sufficient statistics for unknown parameters $\omega = (\beta, \sigma_1^2, \sigma_2^2)$ are given by $\hat\beta_1$, $\hat\beta_2$, S_1 and S_2 where

$$\hat\beta_i = (X_i'X_i)^{-1}X_i'y_i,$$
$$S_i = ||y_i - X_i\hat\beta_i||^2, \quad i = 1, 2,$$

for the Euclidean norm $||u|| = (u'u)^{1/2}$. The common regression coefficients β is estimated based on these statistics. The minimal sufficient statistics are, however, not complete, so that we could not construct uniformly minimum variance unbiased estimators (UMVUE) through the Rao-Blackwell theorem. This demonstrates some difficulty in estimation of the common regression coefficients.

When σ_1^2 and σ_2^2 are known, we would estimate β by the generalized least squares (GLS) estimator

$$\hat{\beta}^{GLS} = \left(\frac{1}{\sigma_1^2}X_1'X_1 + \frac{1}{\sigma_2^2}X_2'X_2\right)^{-1}\left(\frac{1}{\sigma_1^2}X_1'X_1\hat{\beta}_1 + \frac{1}{\sigma_2^2}X_2'X_2\hat{\beta}_2\right), \tag{1.1}$$

which is the UMVUE, the maximum likelihood estimator (MLE) and the best linear unbiased estimator (BLUE). Since σ_1^2 and σ_2^2 are both unknown in our model, the *feasible generalized least squares (FGLS) estimators* are considered by substituting estimators of σ_1^2 and σ_2^2 in the GLS estimator $\hat{\beta}^{GLS}$. These are also called two-stage (or estimated) GLS estimators and two-stage Aitken estimators in econometrics (Taylor(1977, 78), Swamy and Mehta(1979), Kariya(1981), Toyooka and Kariya(1986) and Kurata and Kariya(1996)). Substituting unbiased estimators

$$\hat{\sigma}_i^2 = S_i/m_i, \quad m_i = n_i - p,$$

for σ_i^2, $i = 1, 2$, we get an FGLS estimator of the form

$$\hat{\beta}^{FGLS} = \left(\frac{m_1}{S_1}X_1'X_1 + \frac{m_2}{S_2}X_2'X_2\right)^{-1}\left(\frac{m_1}{S_1}X_1'X_1\hat{\beta}_1 + \frac{m_2}{S_2}X_2'X_2\hat{\beta}_2\right). \tag{1.2}$$

Since $\hat{\beta}^{FGLS}$ is a quite natural, random weighted estimator, one would believe the admissibility of $\hat{\beta}^{FGLS}$ among unbiased estimators.

The main purpose of this paper is to establish the inadmissibility of the FGLS estimator $\hat{\beta}^{FGLS}$ within the class of unbiased estimators. The criterion adopted here for comparing estimators is to minimize the covariance matrices of the estimators, that is, for two unbiased estimators $\hat{\beta}^A$ and $\hat{\beta}^B$ of β, we say that $\hat{\beta}^A$ *is better than* $\hat{\beta}^B$ *in the covariance-matrix criterion* if

$$Cov_\omega(\hat{\beta}^A) = E_\omega[(\hat{\beta}^A - \beta)(\hat{\beta}^A - \beta)'] \leq Cov_\omega(\hat{\beta}^B) \tag{1.3}$$

for every unknown ω and the strict inequality holds for some ω, where the inequality in (1.3) means that $Cov_\omega(\hat{\beta}^B) - Cov_\omega(\hat{\beta}^A)$ is non-negative definite. Since unbiased estimation is focused on in this paper, it is reasonable to utilize the above covariance-matrix criterion. When estimation problems are discussed beyond restriction of the unbiasedness in general, mean squared error matrices should be employed as a measure for evaluating estimators. It is also noted that if the superiority of an estimator is shown in the covariance-matrix criterion, then the superiority of it still holds in the mean squared error (MSE) $\mathrm{tr}E_\omega[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = E_\omega[||\hat{\beta}-\beta||^2]$.

The idea for improving on the FGLS estimator $\hat{\beta}^{FGLS}$ is related to Stein (1964), who provided the innovative decision-theoretic result of inadmissibility of a usual variance estimator by incorporating the information contained in a sample mean. In our model, it is noticed that the statistic $(\hat{\beta}_1-\hat{\beta}_2)(\hat{\beta}_1-\hat{\beta}_2)'$ possesses the information on σ_1^2 and σ_2^2. Since the unbiased estimators of σ_i^2's are used in $\hat{\beta}^{FGLS}$ as stated above, we can imagine that $\hat{\beta}^{FGLS}$ will be improved on by making use of variance estimators incorporating the available information in $(\hat{\beta}_1-\hat{\beta}_2)(\hat{\beta}_1-\hat{\beta}_2)'$. In Section 2, we really develop a new decision-theoretic result of inadmissibility of $\hat{\beta}^{FGLS}$. One of improved procedures is a double shrinkage and unbiased estimator of the form

$$\hat{\beta}^{DS} = \left(\frac{1}{\hat{\sigma}_1^{2*}}X_1'X_1 + \frac{1}{\hat{\sigma}_2^{2*}}X_2'X_2\right)^{-1}\left(\frac{1}{\hat{\sigma}_1^{2*}}X_1'X_1\hat{\beta}_1 + \frac{1}{\hat{\sigma}_2^{2*}}X_2'X_2\hat{\beta}_2\right), \tag{1.4}$$

where

$$\hat{\sigma}_1^{2*} = \min\left\{\frac{S_1}{m_1}, \frac{m_2+2}{m_2}\frac{S_1+(\hat{\beta}_1-\hat{\beta}_2)'X_1'X_1(\hat{\beta}_1-\hat{\beta}_2)}{m_1+p-2}\right\},$$

$$\hat{\sigma}_2^{2*} = \min\left\{\frac{S_2}{m_2}, \frac{m_1+2}{m_1}\frac{S_2+(\hat{\beta}_1-\hat{\beta}_2)'X_2'X_2(\hat{\beta}_1-\hat{\beta}_2)}{m_2+p-2}\right\}.$$

The conditions for $\hat{\sigma}_1^{2*} \neq S_1/m_1$ and $\hat{\sigma}_2^{2*} \neq S_2/m_2$ are given by $(p-2)m_2 > 2m_1$ and $(p-2)m_1 > 2m_2$, respectively. These conditions demonstrate that, for instance, if m_1 $(= n_1 - p)$ is a small number with $m_1 < (p-2)m_2/2$, then the estimator $\hat{\sigma}_1^{2*}$ suggests the use of the information in $(\hat{\beta}_1-\hat{\beta}_2)'X_1'X_1(\hat{\beta}_1-\hat{\beta}_2)$ in order to improve accuracy of estimation for σ_1^2. The inadmissibility of $\hat{\beta}^{FGLS}$ given by (1.2) is thus established when $m_1 < (p-2)m_2/2$ or $m_2 < (p-2)m_1/2$ for $p \geq 3$. In the unbalanced case: $m_1 \neq m_2$, the condition is always satisfied for $p \geq 4$ while when $m_1 = m_2$, it is always guaranteed for $p \geq 5$. The proofs of the dominance results of Section 2 are given in Section 3.

Before stating the main results, it is noted that the estimation of common regression coefficients in our model is related to the problem of recovery of interblock information in balanced incomplete block designs, treated by Khatri and Shah (1974), Brown and Cohen (1974) and Bhattacharya (1980), and the problem of estimating the common mean of two different populations, studied by Cohen and Sackrowitz (1974) and Kubokawa (1987) among others. In general, the problem of estimation of common regression coefficients appears in various applicable models such as heteroscedastic linear models and mixed linear (or variance components) models in biostatistics and econometrics, and

FGLS estimators are heavily and widely exploited. They are quite useful in the case of relatively large sample for their asymptotic efficiency. The problem, however, arises when data enough to estimate the error variances are not available. In this case, as indicated by Rao and Subrahmaniam (1971) and Rao (1980), the information included in sample means may be useful so as to get estimators with higher accuracy. Although the estimation issue given in this paper is limited to the simple situation, the obtained results suggest a possibility of constructing estimators with higher efficiency in more general setting.

2. Inadmissibility of the FGLS Estimators

The feasible generalized least squares estimator of the common regression coefficients β with a more general form is given by

$$\hat{\beta}^{FGLS}(c) = \left(\frac{1}{S_1}X_1'X_1 + \frac{c}{S_2}X_2'X_2\right)^{-1}\left(\frac{1}{S_1}X_1'X_1\hat{\beta}_1 + \frac{c}{S_2}X_2'X_2\hat{\beta}_2\right), \qquad (2.1)$$

where c is a positive constant. The usual choice of c is m_2/m_1, for $m_i = n_i - p$, $i = 1, 2$, which corresponds to the fact that the error variances σ_1^2 and σ_2^2 are unbiasedly estimated, and this choice provides $\hat{\beta}^{FGLS}$ given by (1.2). When the variances are estimated by the MLE in each regression equation, the value of c is n_2/n_1. For the query about existence of the optimal c, we provide the following proposition. The proofs of proposition and theorems given in this section are deferred to Section 3.

Proposition 1. *The optimal value of c when $\rho \to 0$ (resp. $\rho \to \infty$) is given by*

$$\overline{C} = \frac{m_2 + 2}{m_1 - 4} \quad \left(\textit{resp. } \underline{C} = \frac{m_2 - 4}{m_1 + 2}\right).$$

When $d > \overline{C}$ (resp. $d < \underline{C}$), $\hat{\beta}^{FGLS}(d)$ is improved on by $\hat{\beta}^{FGLS}(\overline{C})$ (resp. $\hat{\beta}^{FGLS}(\underline{C})$). If $\underline{C} \le d \le \overline{C}$, then there exist no FGLS estimators $\hat{\beta}^{FGLS}(c)$ being better than $\hat{\beta}^{FGLS}(d)$.

Proposition 1 implies that the constant c should be chosen between $\underline{C}$ and $\overline{C}$. On the other hand, it may be requested that the combined estimator $\hat{\beta}^{FGLS}(c)$ has a uniformly smaller covariance matrix than uncombined estimators $\hat{\beta}_1$ and $\hat{\beta}_2$, which is guaranteed by $\overline{C}/2 \le c \le 2\underline{C}$ as presented by Graybill and Deal (1959), Shinozaki (1678) and Swamy and Mehta (1979). Thereby the constant c may be desirable to be chosen as

$$\max(\underline{C}, \overline{C}/2) \le c \le \min(\overline{C}, 2\underline{C}). \qquad (2.2)$$

For $c = m_2/m_1$, this condition is satisfied by $(m_1 - 8)(m_2 - 2) \geq 16$ and $(m_1 - 2)(m_2 - 8) \geq 16$, or equivalently, $(m_1 = 9, m_2 \geq 18)$, $(m_1 \geq 18, m_2 = 9)$ and $(m_1 \geq 10, m_2 \geq 10)$.

Our interest is to investigate whether the FGLS estimator $\hat{\beta}^{FGLS}(c)$ is admissible in the covariance-matrix criterion (1.3). The admissibility of $\hat{\beta}^{FGLS}(c)$ has never been established (Sinha and Mouqadem(1982)) while one believe the admissibility, for the FGLS estimator is quite natural, random weighted least squares estimator. It is helpful to point out that

$$E[(\hat{\beta}_1 - \hat{\beta}_2)(\hat{\beta}_1 - \hat{\beta}_2)'] = \sigma_1^2 (X_1'X_1)^{-1} + \sigma_2^2 (X_2'X_2)^{-1},$$

which means that the statistic $(\hat{\beta}_1 - \hat{\beta}_2)(\hat{\beta}_1 - \hat{\beta}_2)'$ contains the information on σ_1^2 and σ_2^2. The information in this statistic may be available for estimation of the variances while $\hat{\beta}^{FGLS}(c)$ employs the information in S_1 and S_2 only. For improving on $\hat{\beta}^{FGLS}(c)$, we thus consider the estimator

$$\hat{\beta}^{DS}(\phi, \psi) = \left(\frac{1}{S_1} X_1'X_1 + \frac{c\phi}{S_2\psi} X_2'X_2 \right)^{-1} \left(\frac{1}{S_1} X_1'X_1\hat{\beta}_1 + \frac{c\phi}{S_2\psi} X_2'X_2\hat{\beta}_2 \right), \tag{2.3}$$

where

$$\phi = \phi\left(\frac{1}{S_1} (\hat{\beta}_1 - \hat{\beta}_2)' X_1'X_1 (\hat{\beta}_1 - \hat{\beta}_2) \right),$$
$$\psi = \psi\left(\frac{1}{S_2} (\hat{\beta}_1 - \hat{\beta}_2)' X_2'X_2 (\hat{\beta}_1 - \hat{\beta}_2) \right).$$

We denote $\hat{\beta}_1^S(\phi) = \hat{\beta}^{DS}(\phi, 1)$ and $\hat{\beta}_2^S(\psi) = \hat{\beta}^{DS}(1, \psi)$ and call them *Single Shrinkage Estimators* while we call $\hat{\beta}^{DS}(\phi, \psi)$ the *Double Shrinkage Estimator* for $\phi \neq 1$ and $\psi \neq 1$. As shown by Khatri and Shah (1974), Brown and Cohen (1974) and Swamy and Mehta (1979), the estimator $\hat{\beta}^{DS}(\phi, \psi)$ is unbiased. Using the *Integral-Expression-of-Risk-Difference (IERD)* method given by Takeuchi (1991), Kubokawa (1994) and Kubokawa *et al.*(1994,96), we can establish the following theorem concerning the superiority of the single shrinkage estimator $\hat{\beta}_2^S(\psi)$.

Theorem 1. *For $c < (m_2 + p - 2)/(m_1 + 2)$, assume that*

(a) $\psi(w)$ is nondecreasing and $\lim_{w \to \infty} \psi(w) = 1$,

(b) $\psi(w) \geq \min\{1, \psi^*(w)\}$, where

$$\psi^*(w) = \frac{c(m_1 + 2)}{m_2 + p - 2} \frac{\int_0^w x^{(p+2)/2-1}/(1+x)^{(m_2+p)/2-1} dx}{\int_0^w x^{(p+2)/2-1}/(1+x)^{(m_2+p)/2} dx}. \tag{2.4}$$

Then the single shrinkage unbiased estimator $\hat{\beta}_2^S(\psi)$ is better than the FGLS estimator $\hat{\beta}^{FGLS}(c)$ in the covariance-matrix criterion (1.3).

The condition $c < (m_2+p-2)/(m_1+2)$ guarantees that $\psi(w) \neq 1$ with a positive probability. It can be seen that $\psi^*(w)$ is nondecreasing and

$$\lim_{w\to 0}\psi^*(w) = \frac{c(m_1+2)}{m_2+p-2} \quad \text{and} \quad \lim_{w\to\infty}\psi^*(w) = \frac{c(m_1+2)}{m_2-4}.$$

Also note that

$$\begin{aligned}\frac{\int_0^w x^\alpha/(1+x)^\beta dx}{\int_0^w x^\alpha/(1+x)^{\beta+1}dx} &\le \frac{\int_0^w x^\alpha(1+x)dx}{\int_0^w x^\alpha dx} \\ &= 1+\frac{\alpha+1}{\alpha+2}w \\ &\le 1+w,\end{aligned}$$

where we used the well-known fact (for instance, see Bhattacharya (1984)) that if for positive functions $f(x)$, $g(x)$ and $h(x)$, two functions $g(x)/f(x)$ and $h(x)$ are monotone in the opposite directions, then

$$\frac{\int g(x)h(x)dx}{\int f(x)h(x)dx} \le \frac{\int g(x)dx}{\int f(x)dx}. \tag{2.5}$$

These observations imply that the conditions (a) and (b) of Theorem 1 are satisfied by the shrinkage functions

$$\psi_0(w) = \min\{1,\ \psi^*(w)\}, \tag{2.6}$$

$$\psi_1(w) = \min\left\{1,\ \frac{c(m_1+2)}{m_2+p-2}\left(1+\frac{p+2}{p+4}w\right)\right\}, \tag{2.7}$$

$$\psi_2(w) = \min\left\{1,\ \frac{c(m_1+2)}{m_2+p-2}(1+w)\right\}. \tag{2.8}$$

When $c = \underline{C} = (m_2-4)/(m_1+2)$, we have that $\psi_0(w) = \psi^*(w)$, that is, $\hat\beta_2^S(\psi_0)$ is a smooth estimator improving on $\hat\beta^{FGLS}(\underline{C})$. For the usual choice $c = m_2/m_1$, the inadmissibility of $\hat\beta^{FGLS}(m_2/m_1)$ is established if

$$(p-2)m_1 > 2m_2, \tag{2.9}$$

which is satisfied by $(p=3, m_1 > 2m_2)$, $(p=4, m_1 > m_2)$ or $(p \ge 5, m_1 > 2m_2/(p-2))$.

By the symmetry consideration, it can be seen that the single shrinkage estimator $\hat\beta_1^S(\phi)$ dominates $\hat\beta^{FGLS}(c)$ in the covariance-matrix criterion if *for* $c < (m_2+p-2)/(m_1+2)$,

(a) $\phi(w)$ *is nondecreasing and* $\lim_{w\to\infty}\phi(w) = 1$,

(b) $\phi(w) \geq \min\{1, \phi^*(w)\}$, *where*

$$\phi^*(w) = \frac{m_2+2}{c(m_1+p-2)} \frac{\int_0^w x^{(p+2)/2-1}/(1+x)^{(m_1+p)/2-1}dx}{\int_0^w x^{(p+2)/2-1}/(1+x)^{(m_1+p)/2}dx}. \tag{2.10}$$

By the same arguments as below Theorem 1, it can be verified that these conditions are satisfied by the shrinkage functions

$$\phi_0(w) = \min\{1,\ \phi^*(w)\}, \tag{2.11}$$

$$\phi_1(w) = \min\left\{1,\ \frac{m_2+2}{c(m_1+p-2)}\left(1+\frac{p+2}{p+4}w\right)\right\}, \tag{2.12}$$

$$\phi_2(w) = \min\left\{1,\ \frac{m_2+2}{c(m_1+p-2)}(1+w)\right\}. \tag{2.13}$$

For $c = m_2/m_1$, the sufficient condition for $\hat{\beta}_1^S(\phi)$ to dominate $\hat{\beta}^{FGLS}(m_2/m_1)$ is given by

$$(p-2)m_2 > 2m_1, \tag{2.14}$$

and together with (2.9), it follows that $\hat{\beta}^{FGLS}(m_2/m_1)$ is *inadmissible* if

$m_1 > 2m_2$ or $m_2 > 2m_1$ for $p = 3$,

$m_1 > m_2$ or $m_2 > m_1$ for $p = 4$,

$m_1 > 2m_2/(p-2)$ or $m_2 > 2m_1/(p-2)$ for $p \geq 5$.

In the unbalanced case, that is, $m_1 \neq m_2$, the condition for the inadmissibility of $\hat{\beta}^{FGLS}(m_2/m_1)$ is always satisfied for $p \geq 4$ while when $m_1 = m_2$, the condition is always guaranteed for $p \geq 5$.

We now address the problem of investigating whether the single shrinkage estimators $\hat{\beta}_1^S(\phi)$ and $\hat{\beta}_2^S(\psi)$ can be further improved on by the double shrinkage estimator $\hat{\beta}^{DS}(\phi, \psi)$. This problem has somewhat of technical difficulty in two respects: (1) the shrinkage functions $\phi(w)$ and $\psi(w)$ hold the statistic $(\hat{\beta}_1 - \hat{\beta}_2)(\hat{\beta}_1 - \hat{\beta}_2)'$ in common and (2) $\phi(w)/\psi(w)$ shrinks towards the opposite directions. Under such difficulty, applying the IERD method establishes the following theorem.

Theorem 2. *For* $c > (m_2+2)/(m_1+p-2)$, *assume that*

(a) $\phi(w)$ *is nondecreasing and* $\lim_{w\to\infty}\phi(w) = 1$,

(b) $\phi(w) \geq \phi_2(w)$ *for function* $\phi_2(w)$ *given by (2.13).*

When $\psi(w)/w$ *is nonincreasing in* w, *the condition (b) is replaced with*

(b′) $\phi(w) \geq \min\{1, \phi^{**}(w)\}$, *where*

$$\phi^{**}(w) = \frac{m_2+2}{c(m_1+p-2)} \frac{\int_0^w x^{(p+2)/2}/(1+x)^{(m_2+p)/2}dx}{\int_0^w x^{(p+2)/2}/(1+x)^{(m_2+p)/2+1}dx}. \tag{2.15}$$

Then the double shrinkage estimator $\hat{\beta}^{DS}(\phi, \psi)$ improves on the single shrinkage estimator $\hat{\beta}_2^S(\psi)$ in the covariance-matrix criterion (1.3).

By using the inequality (2.5), it is easily shown that $\phi^*(w) \leq \phi^{**}(w)$ for $\phi^*(w)$ given by (2.10), so that the condition (b$'$) is somewhat restrictive. This restriction is caused by the reason that ϕ and ψ contain the common statistic.

The symmetry consideration can give a similar condition for $\hat{\beta}^{DS}(\phi, \psi)$ to dominate $\hat{\beta}_1^S(\phi)$, and we get a sufficient condition for $\hat{\beta}^{DS}(\phi_2, \psi_2)$ being better than $\hat{\beta}_1^S(\phi_2)$ and $\hat{\beta}_2^S(\psi_2)$.

Corollary. *Assume that*

$$\frac{m_2+2}{m_1+p-2} < c < \frac{m_2+p-2}{m_1+2}. \tag{2.16}$$

Then the double shrinkage estimator $\hat{\beta}^{DS}(\phi_2, \psi_2)$ dominates single shrinkage estimators $\hat{\beta}_1^S(\phi_2)$ and $\hat{\beta}_2^S(\psi_2)$, being better than $\hat{\beta}^{FGLS}(c)$ in the covariance-matrix criterion (1.3), where ϕ_2 and ψ_2 are given by (2.13) and (2.8) respectively.

When $c = m_2/m_1$, the condition (2.16) is satisfied by

$$\frac{2}{p-2}m_2 < m_1 < \frac{p-2}{2}m_2 \quad \text{for} \quad p \geq 5,$$

and then, the improved single shrinkage estimators are further dominated by the double shrinkage estimator $\hat{\beta}^{DS}$ given by (1.4), while $\hat{\beta}^{DS}$ is superior to $\hat{\beta}^{FGLS}$, by (1.2), for $2m_2 < (p-2)m_1$ or $2m_1 < (p-2)m_2$.

We conclude this section with providing the results of Monte Carlo simulation for the relative covariance-matrix improvement

$$100 \times Cov(\hat{\beta}^{FGLS})^{-1/2}\left\{Cov(\hat{\beta}^{FGLS}) - Cov(\hat{\beta})\right\}Cov(\hat{\beta}^{FGLS})^{-1/2}$$

for the single shrinkage estimator $\hat{\beta} = \hat{\beta}_1^S(\phi_2) = \hat{\beta}_1^S$ and the double shrinkage estimator $\hat{\beta} = \hat{\beta}^{DS}(\phi_2, \psi_2) = \hat{\beta}^{DS}$. These are done in the cases where $p = 20$, $(m_1, m_2) = (2, 2)$, $(10, 10)$, $(15, 15)$, $(2, 20)$, $(20, 2)$, $(5, 15)$ and $(15, 5)$ and $\rho = \sigma_2^2/\sigma_1^2 = 0.1$, 0.2, 0.5, 0.75, 1.0, 1.33, 2.0, 5.0 and 10.0. The case where $X_1'X_1 = X_2'X_2$ is only treated for simplicity, and then the relative covariance-matrix improvement is a diagonal matrix with the same diagonal element which is just the relative variance improvement for each component of the estimators. Table 1 reports the average values of this relative variance improvement based on 50,000 replications. From the table, we see that the variance gain of $\hat{\beta}^{DS}$ is relatively bigger in the unbalanced cases $m_1 \neq m_2$ than in the balanced cases $m_1 = m_2$.

Also for larger m_1 and smaller m_2, the variance gains of the single shrinkage estimator $\hat{\beta}_1^S$ are quite small while $\hat{\beta}^{DS}$ holds the reasonable variance gains. This demonstrates that the double shrinkage estimator works effectively in comparison with the single shrinkage one. When $m_1 = m_2 = 2$, the variance improvement of $\hat{\beta}^{DS}$ attains the largest value near $\rho = 1.0$ and approaches zero when ρ tends to zero or infinity. When $m_1 = m_2 = 10$ or 15, in contrast with the case of $m_1 = m_2 = 2$, the variance gain is large near $\rho = 0.2$ and 5.0 while it is small at $\rho = 1.0$. This phenomenon seems to be brought about by the fact that the ratio of unbiased estimators of σ_1^2 and σ_2^2 gives a better estimate of ρ near $\rho = 1.0$ for larger m_1 and m_2.

3. Proofs

We begin with reducing the estimation of β to the equivalent one-dimensional problems. Let Q be a $p \times p$ nonsingular matrix such that $X_1'X_1 = QQ'$ and $X_2'X_2 = QD_\lambda Q'$ where $D_\lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$, λ_i's being eigenvalues of $(X_1'X_1)^{-1}X_2'X_2$. Let $x_1 = (x_{11}, \dots, x_{1p})' = Q'\hat{\beta}_1$, $x_2 = (x_{21}, \dots, x_{2p})' = Q'\hat{\beta}_2$ and $\mu = (\mu_1, \dots, \mu_p)' = Q\beta$. Then $x_1 \sim \mathcal{N}_p(\mu, \sigma_1^2 I_p)$, $x_2 \sim \mathcal{N}_p(\mu, \sigma_2^2 I_p)$ and

$$Q'\hat{\beta}^{DS}(\phi, \psi) = \left(\frac{1}{S_1\phi} I_p + \frac{c}{S_2\psi} D_\lambda \right)^{-1} \left(\frac{1}{S_1\phi} x_1 + \frac{c}{S_2\psi} D_\lambda x_2 \right),$$

where

$$\phi = \phi((x_1 - x_2)'(x_1 - x_2)/S_1),$$
$$\psi = \psi((x_1 - x_2)'D_\lambda(x_1 - x_2)/S_2).$$

Noting that $\hat{\beta}^{DS}(\phi, \psi)$ is an unbiased estimator of be, we see that

$$\begin{aligned} &Q'Cov\left(\hat{\beta}^{DS}(\phi, \psi)\right) Q \\ &= E\left[\left(Q'\hat{\beta}^{DS}(\phi, \psi) - Q'\beta \right) \left(Q'\hat{\beta}^{DS}(\phi, \psi) - Q'\beta \right)' \right] \\ &= \mathrm{diag}\left(E\left[\left(\hat{\mu}_i^{DS}(\phi, \psi) - \mu_i \right)^2 \right],\ i = 1, \cdots, p \right), \end{aligned}$$

where

$$\hat{\mu}_i^{DS}(\phi, \psi) = \left(\frac{1}{S_1\phi} + \frac{c\lambda_i}{S_2\psi} \right)^{-1} \left(\frac{1}{S_1\phi} x_{1i} + \frac{c\lambda_i}{S_2\psi} x_{2i} \right). \tag{3.1}$$

This implies that $\hat{\beta}^{DS}(\phi, \psi)$ is better than $\hat{\beta}^{FGLS}(c)$ in the covariance-matrix criterion if and only if for every i, the variance of $\hat{\mu}_i^{DS}(\phi, \psi)$ is uniformly smaller than that of $\hat{\mu}_i^{DS}(1, 1)$.

Without loss of generality, let $i = 1$ and consider the problem of estimating μ_1 relative to the squared error loss. Let us express $\hat{\mu}_1^{DS}(\phi, \psi)$ as $\hat{\mu}_1^{DS}(\phi, \psi) = \Phi x_{11} + (1 - \Phi)x_{21}$ for $\Phi = S_2\psi/(c\lambda_1 S_1\phi + S_2\psi)$. Also let

$$\hat{\mu}_1^{GLS} = \frac{\rho}{1+\rho} x_{11} + \left(1 - \frac{\rho}{1+\rho} \right) x_{21}$$

for $\rho = \sigma_2^2/(\lambda_1\sigma_1^2)$. The variance of $\hat{\mu}_1^{DS}(\phi,\psi)$ is written as

$$\begin{aligned}
&Var_\rho\left(\hat{\mu}_1^{DS}(\phi,\psi)\right)\\
&= E_\rho\left[\left(\hat{\mu}_1^{DS}(\phi,\psi) - \hat{\mu}_1^{GLS} + \hat{\mu}_1^{GLS} - \mu_1\right)^2\right]\\
&= E_\rho\left[\left\{\left(\Phi - \frac{\rho}{1+\rho}\right)(x_{11}-x_{21}) + \frac{\rho}{1+\rho}(x_{11}-x_{21}) + x_{21} - \mu_1\right\}^2\right]\\
&= E_\rho\left[\left(\Phi - \frac{\rho}{1+\rho}\right)^2 (x_{11}-x_{21})^2\right] + Var_\rho\left(\hat{\mu}_1^{GLS}\right) \qquad (3.2)\\
&\quad + 2E_\rho\left[\left(\Phi - \frac{\rho}{1+\rho}\right)(x_{11}-x_{21})\left\{\frac{\rho}{1+\rho}(x_{11}-x_{21}) + x_{21} - \mu_1\right\}\right].
\end{aligned}$$

Note that Φ is a function of S_1, S_2 and $(x_{1i}-x_{2i})^2$, $i=1,\dots,p$, and that the conditional expectation of $x_{21}-\mu_1$ given $x_{11}-x_{21}$ is

$$E_\rho[x_{21}-\mu_1|x_{11}-x_{21}] = -\frac{\rho}{1+\rho}(x_{11}-x_{21}),$$

which shows that the third term of the r.h.s. of the extreme equality in (3.2) vanishes. Hence,

$$Var_\rho\left(\hat{\mu}_1^{DS}(\phi,\psi)\right) - Var_\rho\left(\hat{\mu}_1^{GLS}\right) = E_\rho\left[\left(\Phi - \frac{\rho}{1+\rho}\right)^2 (x_{11}-x_{21})^2\right], \qquad (3.3)$$

which demonstrates the amount of the estimation error arisen from substituting estimators $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ for σ_1^2 and σ_2^2. The estimation problem of μ_1 is thus reduced to that of $\rho/(1+\rho)$ relative to the loss $(\Phi-\rho/(1+\rho))^2(x_{11}-x_{21})^2$.

Let z_1 be a random variable having a χ_3^2-distribution. Also let $v_i = S_i/\sigma_i^2$ for $i=1,2$ and $z_i = (x_{1i}-x_{2i})^2/(\sigma_1^2+\sigma_2^2/\lambda_i)$ for $i=2,\dots,p$. Then,

$$\begin{aligned}
&E_\rho\left[\left(\frac{S_2\psi}{c\lambda_1 S_1\phi + S_2\psi} - \frac{\rho}{1+\rho}\right)^2 (x_{11}-x_{21})^2\right]\\
&= \sigma_1^2(1+\rho)E_\rho\left[\left(\frac{\rho v_2\overline{\psi}}{cv_1\overline{\phi}+\rho v_2\overline{\psi}} - \frac{\rho}{1+\rho}\right)^2\right] \qquad (3.4)\\
&= \frac{\sigma_1^2\rho^2}{1+\rho}R_1(\rho,c,\phi,\psi),
\end{aligned}$$

where for $F = v_2/v_1$,

$$R_1(\rho,c,\phi,\psi) = E_\rho\left[\left(\frac{(1+\rho)F\overline{\psi}}{c\overline{\phi}+\rho F\overline{\psi}} - 1\right)^2\right], \qquad (3.5)$$

$$\overline{\phi} = \phi\left(\sum_{i=1}^{p}\left(1+\rho\frac{\lambda_1}{\lambda_i}\right)\frac{z_i}{v_1}\right),$$

$$\overline{\psi} = \psi\left(\sum_{i=1}^{p}\left(1+\frac{\lambda_i}{\rho\lambda_1}\right)\frac{z_i}{v_2}\right).$$

The quantity (3.4) is also represented in another form as

$$E_\rho\left[\left(\frac{c\lambda_1 S_1\phi}{c\lambda_1 S_1\phi + S_2\psi} - \frac{1}{1+\rho}\right)^2 (x_{11} - x_{21})^2\right] = \frac{\sigma_1^2}{1+\rho} R_2(\rho, c, \phi, \psi),$$

where for $G = v_1/v_2 = 1/F$,

$$R_2(\rho, c, \phi, \psi) = E_\rho\left[\left(\frac{(1+\rho)cG\overline{\phi}}{cG\overline{\phi} + \rho\overline{\psi}} - 1\right)^2\right], \tag{3.6}$$

and both expressions will be used in the proofs.

We now prove the results given in the previous section. The following lemmas are useful for our purpose.

Lemma 1. *Let $h(x)$ be a nondecreasing and positive function on interval (a, b). If for function $K(x)$ on (a, b), there exists a point x_0 on (a, b) such that $K(x) < 0$ for $x < x_0$ and $K(x) > 0$ for $x > x_0$, then*

$$\int_a^b K(x)h(x)dx \geq h(x_0)\int_a^b K(x)dx,$$

where the equality holds if and only if $h(x)$ is a constant almost everywhere.

Lemma 2. *Let X be a positive random variable with $E[X^{-r-1}] < \infty$ for $r \geq -1$. Then for $0 < \theta < 1$,*

$$\frac{E[(\theta + (1-\theta)X)^{-r}]}{E[(\theta + (1-\theta)X)^{-r-1}]} \geq \min\left\{1, \frac{E[X^{-r}]}{E[X^{-r-1}]}\right\}.$$

Lemma 1 was used in the proof given by Strawderman (1974) and Lemma 2 can be easily verified by the same arguments as in Battacharya (1984).

Proof of Proposition 1. Taking $\rho \to 0$ in the expression (3.5) gives that

$$\lim_{\rho\to 0} R_1(\rho, c, 1, 1) = E[(F/c - 1)^2],$$

which is minimized at $c = E[F^2]/E[F] = (m_2 + 2)/(m_1 - 4) = \overline{C}$. Similarly,

$$\lim_{\rho\to\infty} R_2(\rho, c, 1, 1) = E[(cG - 1)^2],$$

being minimized at $c = E[G]/E[G^2] = (m_2 - 4)/(m_1 + 2) = \underline{C}$.

We next compare two FGLS estimators $\hat{\beta}^{FGLS}(d)$ and $\hat{\beta}^{FGLS}(c)$ for $d < \underline{C}$ or $d > \overline{C}$. For this issue, it is sufficient to consider the difference

$$\begin{aligned}\Delta(\rho, d, c) &= E\left[\left(\frac{(1+\rho)F}{d+\rho F} - 1\right)^2 - \left(\frac{(1+\rho)F}{c+\rho F} - 1\right)^2\right] \qquad (3.7)\\ &= (1+\rho)E\left[\frac{(c-d)F}{(d+\rho F)(c+\rho F)}\left\{\frac{(1+\rho)F}{d+\rho F} + \frac{(1+\rho)F}{c+\rho F} - 2\right\}\right],\end{aligned}$$

which is greater than or equal to

$$\begin{aligned}&2(1+\rho)E\left[\frac{(c-d)F}{(d+\rho F)(c+\rho F)}\left\{\frac{(1+\rho)F}{c+\rho F} - 1\right\}\right] \qquad (3.8)\\ &= 2(1+\rho)E\left[\frac{(c-d)F(F-c)}{(d+\rho F)(c+\rho F)^2}\right].\end{aligned}$$

In the case of $d > \overline{C}$, put $c = \overline{C}$. Then from Lemma 1, we can show that

$$\begin{aligned}E\left[\frac{(\overline{C}-d)F(F-\overline{C})}{(d+\rho F)(\overline{C}+\rho F)^2}\right] &\geq E\left[\frac{\overline{C}-d}{(d+\rho F)(\overline{C}+\rho F)^2}\right]\cdot E\left[F(F-\overline{C})\right]\\ &= 0, \qquad (3.9)\end{aligned}$$

which implies that $\hat{\beta}^{FGLS}(d)$ for $d > \overline{C}$ is dominated by $\hat{\beta}^{FGLS}(\overline{C})$. In the case of $d < \underline{C}$, from (3.7) and (3.8), the same argument with putting $c = \underline{C}$ gives that

$$\begin{aligned}E\left[\frac{(\underline{C}-d)F(F-\underline{C})}{(d+\rho F)(\underline{C}+\rho F)^2}\right] &= E\left[\frac{(\underline{C}-d)G(1-\underline{C}G)}{(dG+\rho)(\underline{C}G+\rho)^2}\right]\\ &\geq E\left[\frac{\underline{C}-d}{(dG+\rho)(\underline{C}G+\rho)^2}\right]\cdot E\left[G(1-\underline{C}G)\right]\\ &= 0,\end{aligned}$$

so that $\hat{\beta}^{FGLS}(d)$ for $d < \underline{C}$ is improved on by $\hat{\beta}^{FGLS}(\underline{C})$.

We shall verify that for $\underline{C} \leq d \leq \overline{C}$, $\hat{\beta}^{FGLS}(d)$ is never dominated by $\hat{\beta}^{FGLS}(c)$. It is now supposed that there exists an estimator $\hat{\beta}^{FGLS}(c)$ such that $\hat{\beta}^{FGLS}(c)$ is better than $\hat{\beta}^{FGLS}(d)$ for $\underline{C} \leq d \leq \overline{C}$. If $d > c$, then we have that $0 \leq \Delta(\rho, d, c)$ for any $\rho > 0$, while from (3.7),

$$\lim_{\rho\to 0}\Delta(\rho, d, c) = \frac{c-d}{dc}E[F]\left\{\frac{\overline{C}}{d} + \frac{\overline{C}}{c} - 2\right\} < 0,$$

which yields the contradiction. If $d < c$, on the other hand, using the expression (3.6) gives that

$$\begin{aligned}0 &\leq \lim_{\rho\to\infty}\left\{R_2(\rho, d, 1, 1) - R_2(\rho, c, 1, 1)\right\}\\ &= (d-c)E\left[G^2\right](d+c-2\underline{C}) < 0,\end{aligned}$$

also yielding the contradiction. Hence Proposition 1 is proved.

Proof of Theorem 1. Using the expression (3.6), we see that it suffices to show that $\Delta_1(\rho) \geq 0$ for every $\rho > 0$, where

$$\Delta_1(\rho) = E_\rho\left[\left(\frac{(1+\rho)cG}{cG+\rho}-1\right)^2 - \left(\frac{(1+\rho)cG}{cG+\rho\overline{\psi}}-1\right)^2\right]. \tag{3.10}$$

For simplicity, let $\gamma_{2i} = 1 + \lambda_i/(\rho\lambda_1)$ for $i = 1, \dots, p$ and let $u_2 = \sum_{i=1}^p \gamma_{2i} z_i / v_2$. From the condition (a), note that $\lim_{t\to\infty}\psi(tu_2) = 1$. Applying the Integral-Expression-of-Risk-Difference (IERD) method to the difference $\Delta_1(\rho)$ gives that

$$\begin{aligned}
\Delta_1(\rho) &= E_\rho\left[\left(\frac{(1+\rho)cG}{cG+\rho\psi(tu_2)}-1\right)^2\Big|_{t=1}^{\infty}\right] \\
&= E_\rho\left[\int_1^\infty \frac{d}{dt}\left\{\left(\frac{(1+\rho)cG}{cG+\rho\psi(tu_2)}-1\right)^2\right\}dt\right] \\
&= -2E_\rho\left[\int_1^\infty \left(\frac{(1+\rho)cG}{cG+\rho\psi(tu_2)}-1\right)\frac{(1+\rho)cG\rho u_2\psi'(tu_2)}{\{(cG+\rho\psi(tu_2)\}^2}dt\right] \\
&= -2\rho(1+\rho)cE^{v_1,v_2}\left[\int\cdots\int\int_1^\infty\left(\frac{(1+\rho)cG}{cG+\rho\psi(\sum_i\gamma_{2i}z_it/v_2)}-1\right)\right. \\
&\quad \left.\times\frac{G(\sum_i\gamma_{2i}z_it/v_2)\psi'(\sum_i\gamma_{2i}z_it/v_2)}{\{cG+\rho\psi(\sum_i\gamma_{2i}z_it/v_2)\}^2}dtf_3(z_1)\prod_{i=2}^p f_1(z_i)\prod_{i=1}^p dz_i\right],
\end{aligned} \tag{3.11}$$

where $f_k(z)$ designates a density function of a χ_k^2-distribution. Making the transformations $w_i = z_it/v_2$ with $dw_i = (t/v_2)dz_i$ for $i = 1, \dots, p$, we observe that

$$\begin{aligned}
\frac{\Delta_1(\rho)}{2\rho(1+\rho)c} = -E_\rho^{v_1,v_2}&\left[\int\cdots\int\int_1^\infty\left(\frac{(1+\rho)cG}{cG+\rho\psi(\sum_i\gamma_{2i}w_i)}-1\right)\right. \\
&\times\frac{G(\sum_i\gamma_{2i}w_i)\psi'(\sum_i\gamma_{2i}w_i)}{\{cG+\rho\psi(\sum_i\gamma_{2i}w_i)\}^2} \\
&\left.\times\frac{v_2}{t^2}f_3\left(\frac{v_2}{t}w_1\right)\prod_{i=2}^p\left\{\frac{v_2}{t}f_1\left(\frac{v_2}{t}w_i\right)\right\}dt\prod_{i=1}^p dw_i\right].
\end{aligned} \tag{3.12}$$

Since $\psi(u_1)$ is nondecreasing, we have $\Delta_1(\rho) \geq 0$ for every $\rho > 0$ if

$$\begin{aligned}
E^{v_1,v_2}&\left[\left(\frac{(1+\rho)cG}{cG+\rho\psi(\sum_i\gamma_{2i}w_i)}-1\right)\frac{G}{\{cG+\rho\psi(\sum_i\gamma_{2i}w_i)\}^2}\right. \\
&\left.\times\int_1^\infty\frac{v_2}{t^2}f_3\left(\frac{v_2}{t}w_1\right)\prod_{i=2}^p\left\{\frac{v_2}{t}f_1\left(\frac{v_2}{t}w_i\right)\right\}dt\right] \leq 0,
\end{aligned} \tag{3.13}$$

for every $\rho > 0$ and every $w_i > 0$, $i = 1, \ldots, p$. Here,

$$\frac{v_2}{t^2} f_3\left(\frac{v_2}{t} w_1\right) \prod_{i=2}^{p} \left\{\frac{v_2}{t} f_1\left(\frac{v_2}{t} w_i\right)\right\} \text{ is proportional to } \frac{1}{t}\left(\frac{v_2}{t}\right)^{(p+4)/2-1} e^{-v_2 \sum w_i/t}.$$

By making the transformation $y = v_2 \sum w_i/t$ with $dy = (v_2 \sum w_i/t^2)dt$, the condition (3.13) is written by

$$E^{v_1,v_2}\left[\left(\frac{(1+\rho)cG}{cG+\rho\psi} - 1\right)\frac{G}{(cG+\rho\psi)^2} H_2\left(v_2 \sum_i w_i\right)\right] \leq 0,$$

or

$$\frac{E^{v_1,v_2}[\{\theta + (1-\theta)\psi/(cG)\}^{-2} G^{-1} H_2(v_2 \sum_i w_i)]}{E^{v_1,v_2}[\{\theta + (1-\theta)\psi/(cG)\}^{-3} G^{-1} H_2(v_2 \sum_i w_i)]} \geq 1 \tag{3.14}$$

for every ρ and w_i's, where $\theta = 1/(1+\rho)$ and

$$H_2\left(v_2 \sum_i w_i\right) = \int_0^{v_2 \Sigma_i w_i} y^{(p+2)/2-1} e^{-y/2} dy.$$

Let $E^*[\cdot]$ stand for the expectation with respect to the probability measure

$$P^*\{(v_1, v_2) \in A\} = E^{v_1,v_2}\left[I_A G^{-1} H_2\left(v_2 \sum_i w_i\right)\right] / E^{v_1,v_2}\left[G^{-1} H_2\left(v_2 \sum_i w_i\right)\right].$$

By applying Lemma 2 to the l.h.s. of (3.14), it follows that

$$\frac{E^*[\{\theta + (1-\theta)\psi/(cG)\}^{-2}]}{E^*[\{\theta + (1-\theta)\psi/(cG)\}^{-3}]} \geq \min\left\{1, \frac{E^*[(cG/\psi)^2]}{E^*[(cG/\psi)^3]}\right\}, \tag{3.15}$$

which is greater than or equal to one if

$$\begin{aligned}
\psi\left(\sum_i \gamma_{2i} w_i\right) &\geq c\frac{E^*[G^3]}{E^*[G^2]} \\
&= c\frac{E^{v_1,v_2}[(v_1/v_2)^2 H_2(v_2 \sum_i w_i)]}{E^{v_1,v_2}[(v_1/v_2) H_2(v_2 \sum_i w_i)]} \\
&= c(m_1+2)\frac{E[v_2^{-2} H_2(v_2 \sum_i w_i)]}{E[v_2^{-1} H_2(v_2 \sum_i w_i)]} \\
&= \frac{c(m_1+2)}{m_2+p-2} \frac{\int_0^{\Sigma w_i} x^{(p+2)/2-1}/(1+x)^{(m_2+p)/2-1} dx}{\int_0^{\Sigma w_i} x^{(p+2)/2-1}/(1+x)^{(m_2+p)/2} dx},
\end{aligned} \tag{3.16}$$

for every ρ and w_i's. Since ψ is nondecreasing, $\psi(\sum_i (1 + \lambda_i/(\rho\lambda_1))w_i) \geq \psi(\sum_i w_i)$. Therefore the inequality (3.16) is satisfied by the condition (b) of Theorem 1, which is established.

Proof of Theorem 2. The proof is done by the similar arguments as in the proof of Theorem 1 except that two functions ϕ and ψ include the common statistic. From the expression (3.5), it is sufficient to show that $\Delta_2(\rho) \geq 0$ for every $\rho > 0$, where

$$\Delta_2(\rho) = E_\rho \left[\left(\frac{(1+\rho)F\overline{\psi}}{c+\rho F\overline{\psi}} - 1 \right)^2 - \left(\frac{(1+\rho)F\overline{\psi}}{c\overline{\phi}+\rho F\overline{\psi}} - 1 \right)^2 \right]. \tag{3.17}$$

Let $\gamma_{1i} = 1 + \rho\lambda_1/\lambda_i$, $i = 1, \dots, p$, and $u_1 = \sum_{i=1}^{p} \gamma_{1i} z_i / v_1$. Using the IERD method, for γ_{2i} and u_2 defined in the proof of Theorem 1, we observe that

$$\begin{aligned}
\Delta_2(\rho) &= E_\rho \left[\int_1^\infty \frac{d}{dt} \left\{ \left(\frac{(1+\rho)F\psi(u_2)}{c\phi(tu_1) + \rho F\psi(u_2)} - 1 \right)^2 \right\} dt \right] \\
&= -2E_\rho \left[\int_1^\infty \left(\frac{(1+\rho)F\psi(u_2)}{c\phi(tu_1)+\rho F\psi(u_2)} - 1 \right) \frac{(1+\rho)F\psi(u_2)cu_1\phi'(tu_1)}{\{c\phi(tu_1)+\rho F\psi(u_2)\}^2} dt \right] \\
&= -2(1+\rho)cE_\rho \left[\int_1^\infty \left(\frac{(1+\rho)F\psi(\sum_i \gamma_{2i} z_i/v_2)}{c\phi(\sum_i \gamma_{1i} z_i t/v_1) + \rho F\psi(\sum_i \gamma_{2i} z_i / v_2)} - 1 \right) \right. \\
&\quad \left. \times \frac{F(\sum_i \gamma_{1i} z_i/v_1)\psi(\sum_i \gamma_{2i} z_i/v_2)\phi'(\sum_i \gamma_{1i} z_i t/v_1)}{\{c\phi(\sum_i \gamma_{1i} z_i t/v_1) + \rho F\psi(\sum_i \gamma_{2i} z_i/v_2)\}^2} dt \right].
\end{aligned} \tag{3.18}$$

Making the transformations $w_i = (t/v_1)z_i$ with $dw_i = (t/v_1)dz_i$ gives that

$$\begin{aligned}
\Delta_2(\rho) = -2(1+\rho)cE^{v_1,v_2} &\left[\int \cdots \int \int_1^\infty \left(\frac{(1+\rho)F\psi(\sum_i \gamma_{2i} w_i/(tF))}{c\phi(\sum_i \gamma_{1i} w_i) + \rho F\psi(\sum_i \gamma_{2i} w_i/(tF))} - 1 \right) \right. \\
&\times \frac{F\{\sum_i \gamma_{1i} w_i\}\psi(\sum_i \gamma_{2i} w_i/(tF))\phi'(\sum_i \gamma_{1i} w_i)}{\{c\phi(\sum_i \gamma_{1i} w_i) + \rho F\psi(\sum_i \gamma_{2i} w_i/(tF))\}^2} \\
&\left. \times \frac{v_1}{t^2} f_3\left(\frac{v_1}{t} w_1\right) \prod_{i=2}^{p} \left\{ \frac{v_1}{t} f_1\left(\frac{v_1}{t} w_i\right) \right\} dt \prod_{i=1}^{p} dw_i \right].
\end{aligned} \tag{3.19}$$

Making the transformation $y = v_1 \sum_i w_i / t$ again and using the same arguments as in the proof of Theorem 1, we can see that $\Delta_2(\rho) \geq 0$ for every $\rho > 0$ if

$$\begin{aligned}
E^{v_1,v_2} &\left[\int_0^{v_1 \Sigma w_i} \left(\frac{(1+\rho)F\psi(d_2 y/v_2)}{c\phi(d_1) + \rho F\psi(d_2 y/v_2)} - 1 \right) \right. \\
&\quad \left. \times \frac{d_1 F\psi(d_2 y/v_2)\phi'(d_1)}{\{c\phi(d_1) + \rho F\psi(d_2 y/v_2)\}^2} y^{(p+2)/2-1} e^{-y/2} dy \right] \leq 0,
\end{aligned} \tag{3.20}$$

for every $\rho > 0$ and every $w_i > 0$, where $d_1 = \sum_i \gamma_{1i} w_i$ and $d_2 = \sum_i \gamma_{2i} w_i / \sum_i w_i$. Since $\phi'(d_1) \geq 0$, the inequality (3.20) is equivalent to the condition that

$$\frac{E^{y,v_1,v_2}[\{\theta c\phi/(F\psi) + 1 - \theta\}^{-2} g(y, v_1, v_2)]}{E^{y,v_1,v_2}[\{\theta c\phi/(F\psi) + 1 - \theta\}^{-3} g(y, v_1, v_2)]} \leq 1 \tag{3.21}$$

for every ρ and w_i's, where

$$g(y, v_1, v_2) = \frac{1}{F\psi(d_2 y/v_2)} y^{(p+2)/2-1} e^{-y/2} I(y \le v_1 \sum_i w_i),$$

for indicator function $I(\cdot)$. By the similar arguments as in (3.14), (3.15) and (3.16), it is sufficient to show that $E[(F\psi/c\phi)^2 g(y, v_1, v_2)]/E[(F\psi/c\phi)^3 g(y, v_1, v_2)] \ge 1$, or

$$c\phi(\sum_i \gamma_{1i} w_i) \ge \frac{E^{v_1,v_2}[\int_0^{v_1 \Sigma w_i} (v_2/v_1)^2 \psi^2(d_2 y/v_2) y^{(p+2)/2-1} e^{-y/2} dy]}{E^{v_1,v_2}[\int_0^{v_1 \Sigma w_i} (v_2/v_1) \psi(d_2 y/v_2) y^{(p+2)/2-1} e^{-y/2} dy]}. \tag{3.22}$$

Note that $\phi(\sum_i (1 + \rho\lambda_1/\lambda_i) w_i) \ge \phi(\sum_i w_i)$ and that $\psi(d_2 y/v_2) \le 1$. Hence the inequality (3.22) is satisfied if for $w = \sum_i w_i$,

$$c\phi(w) \ge \frac{E^{v_1,v_2}[\int_0^{v_1 w} (v_2/v_1)^2 \psi(d_2 y/v_2) y^{(p+2)/2-1} e^{-y/2} dy]}{E^{v_1,v_2}[\int_0^{v_1 w} (v_2/v_1) \psi(d_2 y/v_2) y^{(p+2)/2-1} e^{-y/2} dy]}, \tag{3.23}$$

or

$$\begin{aligned}
&\int_0^w E^{v_1,v_2} \left[\left(c\phi - \frac{v_2}{v_1} \right) \frac{v_2}{v_1} \psi \left(d_2 \frac{v_1}{v_2} x \right) v_1 (v_1 x)^{(p+2)/2-1} e^{-v_1 x/2} \right] dx \\
&= (const.) \int_0^w E^{z,v_2} \left[\left(c\phi - (1+x)\frac{v_2}{z} \right) \frac{v_2}{z} \psi \left(d_2 \frac{x}{1+x} \frac{z}{v_2} \right) \right] h_1(x) dx \qquad (3.24) \\
&\ge 0,
\end{aligned}$$

where $h_1(x) = x^{(p+2)/2-1}/(1+x)^{(m_1+p)/2}$, and z is a random variable having a $\chi^2_{m_1+p+2}$-distribution. By applying Lemma 1 to the integrand in (3.24) with respect to the random variable v_2/z, it is evaluated by

$$\begin{aligned}
&\int_0^w E^{z,v_2} \left[\left(c\phi - (1+x)\frac{v_2}{z} \right) \frac{v_2}{z} \psi \left(d_2 \frac{x}{1+x} \frac{z}{v_2} \right) \right] h_1(x) dx \qquad (3.25) \\
&\ge \int_0^w E^{z,v_2} \left[\psi \left(d_2 \frac{x}{1+x} \frac{z}{v_2} \right) \right] \cdot E^{z,v_2} \left[\left(c\phi - (1+x)\frac{v_2}{z} \right) \frac{v_2}{z} \right] h_1(x) dx.
\end{aligned}$$

From the r.h.s. of (3.25), we get one sufficient condition that

$$\begin{aligned}
c\phi(w) &\ge (1+w) E[(v_2/z)^2]/E[v_2/z] \qquad (3.26) \\
&= (1+w)(m_2+2)/(m_1+p-2),
\end{aligned}$$

which is guaranteed by the condition (b). When one can impose the condition that $\psi(w)/w$ is nonincreasing, by applying Lemma 1, the r.h.s. of (3.25) is evaluated by

$$\begin{aligned}
&\int_0^w E^{z,v_2} \left[\psi \left(d_2 \frac{x}{1+x} \frac{z}{v_2} \right) \frac{1+x}{x} \right] \cdot E^{z,v_2} \left[\left(c\phi - (1+x)\frac{v_2}{z} \right) \frac{v_2}{z} \right] \left\{ \frac{x}{1+x} h_1(x) \right\} dx \\
&\ge (const.) \int_0^w E^{z,v_2} \left[\psi \left(d_2 \frac{x}{1+x} \frac{z}{v_2} \right) \frac{1+x}{x} \right] \left\{ \frac{x}{1+x} h_1(x) \right\} dx \qquad (3.27) \\
&\times \int_0^w E^{z,v_2} \left[\left(c\phi - (1+x)\frac{v_2}{z} \right) \frac{v_2}{z} \right] \left\{ \frac{x}{1+x} h_1(x) \right\} dx.
\end{aligned}$$

The r.h.s. of (3.27) is nonnegative if

$$c\phi(w) \geq \frac{\int_0^w E^{z,v_2}[(v_2/z)^2]xh_1(x)dx}{\int_0^w E^{z,v_2}[v_2/z]\{x/(1+x)\}h_1(x)dx}$$

$$= \frac{m_2+2}{m_1+p-2} \frac{\int_0^w x^{(p+2)/2}/(1+x)^{(m_1+p)/2}dx}{\int_0^w x^{(p+2)/2}/(1+x)^{(m_1+p)/2+1}dx},$$

which is guaranteed by the condition (b′). Therefore the proof of Theorem 2 is complete.

Acknowledgements. This research was supported in part by a grant from the Shirakawa Institute of Animal Genetics and the Japan Livestock Technology Association, by the Ministry of Education, Japan, Grant No. 08780216 and by a grant from the Research Institute for the Japanese Economy, University of Tokyo. I am also grateful to Mr. M. Ushijima for his help in the simulation experiments.

REFERENCES

Bhattacharya, C.G.(1980). Estimation of a common mean and recovery of interblock information. *Ann. Statist.* **8**, 205-211.

Bhattacharya, C.G.(1984). Two inequalities with an application. *Ann. Inst. Statist. Math.* **36**, 129-134.

Brown, L.D. and Cohen, A.(1974). Point and confidence estimation of a common mean and recovery of interblock information. *Ann. Statist.* **2**, 963-976.

Cohen, A. and Sackrowitz, H.B.(1974). On estimating the common mean of two normal distributions. *Ann. Statist.* **2**, 1274-1282.

Graybill, F.A. and Deal, R.B.(1959). Combining unbiased estimators. *Biometrics* **15**, 543-550.

Kariya, T.(1981). Bounds for the covariance matrices of Zellner's estimator in the SUR model and the 2SAE in a heteroscedastic model. *J. Amer. Statist. Assoc.* **76**, 975-979.

Khatri, C.G. and Shah, K.R.(1974). Estimation of location parameters from two linear models under normality. *Comm. Statist. A - Theory Methods* **3**, 647-663.

Kubokawa, T.(1987). Admissible minimax estimation of a common mean of two normal populations. *Ann. Statist.* **15**, 1245-1256.

Kubokawa, T.(1994). A unified approach to improving equivariant estimators. *Ann. Statist.* **22**, 290-299.

Kubokawa, T. and Saleh, A.K.Md.E.(1994). Estimation of location and scale parameters under order restrictions. *J. Statist. Research* **28**, 41-51.

Kubokawa, T. and Srivastava, M.S.(1996). Double shrinkage estimators of ratio of variances. In *Multidimensional Statistical Analysis and Theory of Random Matrices* (Proceedings of the Sixth Eugene Lukacs Symposium, Bowling Green, OH, USA, 29-30 March 1996, Editors: A.K. Gupta and V.L. Girko), 139-154, VSP, Netherlands.

Kurata, H. and Kariya, T.(1996). Least upper bound for the covariance matrix of a generalized least squares estimator in regression with applications to a seemingly unrelated regression model and a heteroscedastic model. *Ann. Statist.* **24**, 1547-1559.

Rao, J.N.K.(1980). Estimating the common mean of possibly different normal populations: A simulation study. *J. Amer. Statist. Assoc.*, **75**, 447-453.

Rao, J.N.K. and Subrahmaniam, K.(1971). Combining independent estimators and estimation in linear regression with unequal variances. *Biometrics*, **27**, 971-990.

Shinozaki, N.(1978). A note on estimating the common mean of k normal distributions and Stein problem. *Comm. Statist. A - Theory Methods* **7**, 1421-1432.

Sinha, B.K. and Mouqadem, O.(1982). Estimation of the common mean of two univariate normal populations. *Comm. Statist. A - Theory Methods* **11**, 1603-1614.

Stein, C.(1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.*, **16**, 155-160.

Strawderman, W.E.(1974). Minimax estimation of powers of the variance of a normal population under squared error loss. *Ann. Statist.*, **2**, 190-198.

Swamy, P.A.V.B. and Mehta, J.S.(1979). Estimation of common coefficients in two regression equations. *J. Econometrics* **10**, 1-14.

Takeuchi, K.(1991). Personal Communications.

Taylor, W.E.(1977). Small sample properties of a class of two stage Aitken estimators. *Econometrica* **45**, 497-508.

Taylor, W.E.(1978). The heteroscedastic linear model: Exact finite sample results. *Econometrica* **46**, 663-675.

Toyooka, Y. and Kariya, T.(1986). An approach to upper bound problems for risks of the generalized least squares estimators. *Ann. Statist.* **14**, 679-690.

Table 1. The Relative Variance Improvements of the Single and Double Shrinkage Estimators $\hat{\beta}_1^S = \hat{\beta}_1^S(\phi_2)$ and $\hat{\beta}^{DS} = \hat{\beta}^{DS}(\psi_2, \phi_2)$ for $p = 20$

ρ	0.1	0.2	0.5	0.75	1.0	1.33	2.0	5.0	10.0
$m_1 = 2$ $m_2 = 2$									
$\hat{\beta}_1^S$	0.052	0.114	0.196	0.179	0.142	0.093	0.038	0.004	0.000
$\hat{\beta}^{DS}$	0.052	0.116	0.249	0.291	0.304	0.293	0.251	0.122	0.051
$m_1 = 10$ $m_2 = 10$									
$\hat{\beta}_1^S$	0.591	0.723	0.584	0.364	0.220	0.116	0.034	0.001	0.000
$\hat{\beta}^{DS}$	0.591	0.723	0.623	0.506	0.497	0.554	0.694	0.793	0.658
$m_1 = 15$ $m_2 = 15$									
$\hat{\beta}_1^S$	0.739	0.813	0.582	0.345	0.197	0.090	0.016	0.000	0.000
$\hat{\beta}^{DS}$	0.739	0.813	0.611	0.448	0.399	0.434	0.579	0.814	0.746
$m_1 = 2$ $m_2 = 20$									
$\hat{\beta}_1^S$	0.535	1.204	2.819	3.439	3.576	3.356	2.502	0.359	0.019
$\hat{\beta}^{DS}$	0.535	1.204	2.819	3.439	3.576	3.356	2.502	0.359	0.019
$m_1 = 20$ $m_2 = 2$									
$\hat{\beta}_1^S$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\hat{\beta}^{DS}$	0.034	0.419	2.614	3.539	3.769	3.601	2.941	1.269	0.571
$m_1 = 5$ $m_2 = 15$									
$\hat{\beta}_1^S$	1.040	1.614	2.155	1.897	1.484	0.998	0.409	0.008	0.000
$\hat{\beta}^{DS}$	1.040	1.614	2.156	1.901	1.491	1.009	0.427	0.038	0.029
$m_1 = 15$ $m_2 = 5$									
$\hat{\beta}_1^S$	0.020	0.022	0.012	0.006	0.004	0.002	0.001	0.000	0.000
$\hat{\beta}^{DS}$	0.020	0.042	0.510	1.092	1.588	1.987	2.240	1.700	1.120