96-F-12

Estimation of Variance Components under Kullback-Leibler Loss with Applications

by

Tatsuya Kubokawa
A. K. Md. Ehsane Saleh
Yoshihiko Konno
Masaru Ushijima

June 1996

Estimation of Variance Components under Kullback-Leibler Loss with Applications

Tatsuya Kubokawa, A.K.Md. Ehsane Saleh, Yoshihiko Konno and Masaru Ushijima

University of Tokyo, Carleton University and Chiba University

Estimation of variance components in a mixed linear model with two variance components has been discussed for many years. The problems arise in the incompleteness of the minimal sufficient statistics and the drawback of every unbiased estimator taking a negative value with a positive probability for the 'between' component of variance. This paper addresses to resolve the latter undesirable property. For evaluating estimators, the expected Kullback-Leibler loss function is utilized instead of the usual mean squared error(MSE). The class of estimators improving on the ANOVA(unbiased) estimators by the Henderson method (III) are constructed and out of the class, the positive and useful estimators and the empirical Bayes, the generalized Bayes estimators are derived. Applications of the proposed estimators are given (1) to the generalized least squares(GLS) estimation of the regression coefficients, (2) to the GLS F test for a linear hypothesis and (3) to the two-stage prediction which is related to the small-area estimation. Finally the confidence intervals of the variance components are discussed, and several improvements on the minimum ratio(shortest unbiased) confidence interval related to the Kullback-Leibler loss are proposed.

AMS 1980 subject classifications. Primary 62J10; secondary 62F10, 62F11.

Key words and phrases. Variance components model, unbalanced mixed linear model, point estimation, confidence interval, ANOVA estimator, truncated estimator, Kullback-Leibler information loss, small-area estimation problem, generalized least squares estimation.

1. Introduction and Notations

Consider two-stage cluster sampling in which k clusters are drawn at random at the first stage and n_i elements are drawn at random from the ith sampled cluster at the second stage. $(\Sigma n_i = N)$

Following Fuller and Battese(1973), Wu et al.(1988) and Rao et al.(1993), we consider the nested error regression model

$$y_{ij} = \mathbf{x}'_{ij}\beta + v_i + e_{ij}, \quad j = 1, \ldots, n_i, \; i = 1, \ldots, k \tag{1.1}$$

where $\mathbf{x}_{ij} = (x_{ij0}, x_{ij1}, \ldots, x_{ij,p-1})'$ with $x_{ij0} = 1$ is a vector of known covariates, $\beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$ is a vector of unknown regression coefficients, $v_i \sim \mathcal{N}(0, \sigma_v^2)$, $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$, and v_i's are independent of e_{ij}'s. This model is dealt with in Battese et al.(1988) for prediction of county crop areas (small areas) using survey and satellite data. It is also known as an error component model in econometrics.

For the matrix representation of (1.1), let $\mathbf{y} = (\mathbf{y}'_1, \ldots, \mathbf{y}'_k)'$ with $\mathbf{y}'_i = (y_{i1}, \ldots, y_{i,n_i})$, $\mathbf{X} = (\mathbf{X}'_1 | \cdots | \mathbf{X}'_k)'$ with $\mathbf{X}'_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{i,n_i})$, $\mathbf{v} = (v_1, \ldots, v_k)'$ and $\mathbf{e} = (\mathbf{e}'_1, \ldots, \mathbf{e}'_k)'$ with $\mathbf{e}'_i = (e_{i1}, \ldots, e_{i,n_i})$. Then (1.1) is written as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \quad \mathbf{u} = \mathbf{Z}\mathbf{v} + \mathbf{e} \tag{1.2}$$

where $\mathbf{Z} = \oplus_1^k \mathbf{j}_{n_i}$, the block diagonal form with $n_i \times 1$ unit vector $\mathbf{j}_{n_i}$. This is a mixed model with two variance components and the specific design matrix. Since $\mathbf{v} \sim \mathcal{N}_k(\mathbf{0}, \sigma_v^2 \mathbf{I}_k)$ and $\mathbf{e} \sim \mathcal{N}_N(\mathbf{0}, \sigma_e^2 \mathbf{I}_N)$, $\mathbf{u}$ has $\mathcal{N}_N(\mathbf{0}, (\sigma_v^2 + \sigma_e^2)\mathbf{V}(\rho))$ where

$$\mathbf{V}(\rho) = (1 - \rho)\mathbf{I}_N + \rho \mathbf{Z}\mathbf{Z}' \tag{1.3}$$

with $\rho = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2)$ and $\mathbf{Z}\mathbf{Z}' = \oplus_1^k \mathbf{J}_{n_i}$ for $\mathbf{J}_{n_i} = \mathbf{j}_{n_i}\mathbf{j}'_{n_i}$.

In the statistical inference of the regression coefficient β, useful statistics are functions of ρ in the case where ρ is known. For point estimation of β, the generalized least squares(GLS) estimator is given by

$$\hat{\beta}(\rho) = (\mathbf{X}'\mathbf{V}(\rho)^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}(\rho)^{-1}\mathbf{y}. \tag{1.4}$$

An exact F test treated by Rao et al.(1993) for a linear hypothesis, and the best linear unbiased predictors of some quantity given by Peixoto and Harville(1986) and Battese et al.(1988) are also functions of ρ. Since ρ is unknown in practice, ρ must be estimated based on appropriate estimators $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ of σ_v^2 and σ_e^2. This demonstrates a motivation of estimation of variance components. Since σ_v^2 and σ_e^2 express the extent of dispersion of the first-stage sampling (or the block effects) and the second-stage sampling (or error terms), it may be also important to estimate the variance components themselves.

In this paper, we discuss the point and interval estimation of the variance components with applications to the above problems for the regression coefficients. To derive the estimators of the variance components, following Mathew et al.(1992), we consider the statistics that are invariant under the group of transformation $\mathbf{y} \to \mathbf{y} + \mathbf{Xa}$ where $\mathbf{a} \in \mathbf{R}^p$ is any $p \times 1$ vector. Let $r = \text{rank}(\mathbf{X})$ and let $\mathbf{P}$ be an $N \times (N-r)$ matrix satisfying $\mathbf{P}'\mathbf{X} = \mathbf{0}$ and $\mathbf{P}'\mathbf{P} = \mathbf{I}_{N-r}$. Then

$$\mathbf{P}'\mathbf{y} \sim \mathcal{N}_{N-r}(\mathbf{0}, \sigma_e^2\mathbf{I}_{N-r} + \sigma_v^2\mathbf{P}'\mathbf{ZZ}'\mathbf{P}).$$

Let $\lambda_1 > \lambda_2 > \cdots > \lambda_\ell > 0$ be the distinct non-zero eigenvalues of $\mathbf{P}'\mathbf{ZZ}'\mathbf{P}$ with respect multiplicities $m_1, m_2, \ldots, m_\ell$. In our model (1.1), it is noted that these eigenvalues are distinct n_i's, and that $\text{rank}(\mathbf{P}'\mathbf{ZZ}'\mathbf{P}) = \Sigma m_i = k-1$ since the columns of $\mathbf{Z}$ sum to $\mathbf{1}$ and $\mathbf{P}$ is orthogonal to $\mathbf{1}$. Consider the spectral decomposition

$$\mathbf{P}'\mathbf{ZZ}'\mathbf{P} = \Sigma_{i=1}^{\ell}\lambda_i\mathbf{E}_i$$

where $\mathbf{E}_i$ is an idempotent matrix of rank m_i and $\mathbf{E}_i\mathbf{E}_j = \mathbf{0}$ $(i \neq j)$. Writing $\mathbf{E}_{\ell+1} = \mathbf{I}_{N-r} - \Sigma_{i=1}^{\ell}\mathbf{E}_i$ and assuming $N-r-k+1 > 0$, we see that $m_{\ell+1} = \text{rank}(\mathbf{E}_{\ell+1}) = N-r-k+1 > 0$ and that $\mathbf{y}'\mathbf{y}$ can be decomposed as $\mathbf{y}'\mathbf{y} = \Sigma_{i=1}^{\ell+1}\mathbf{y}'\mathbf{PE}_i\mathbf{P}'\mathbf{y}$ where

$$\begin{aligned} s_i^2 &= \mathbf{y}'\mathbf{PE}_i\mathbf{P}'\mathbf{y} \sim (\sigma_e^2 + \lambda_i\sigma_v^2)\chi_{m_i}^2, \ i = 1, \ldots, \ell \\ s_{\ell+1}^2 &= \mathbf{y}'\mathbf{PE}_{\ell+1}\mathbf{P}'\mathbf{y} \sim \sigma_e^2\chi_{N-r-k+1}^2 \end{aligned} \tag{1.5}$$

and s_i^2's are mutually independent $(i = 1, \ldots, \ell+1)$.

Invariant estimators for σ_e^2 and σ_v^2 are constructed based on s_i^2's, but the uniformly minimum variance unbiased estimator does not exist in the unbalanced case for lack of completeness of the statistics s_i^2's. This leads to production of various kinds of unbiased estimators. For instance, the ANOVA estimators derived by the well-known Henderson (1953) method (III) is expressed in the above notation by

$$\begin{aligned} \hat{\sigma}_e^{2UB} &= \frac{1}{\nu}S_1, \quad \nu = N-r-k+1 \\ \hat{\sigma}_v^{2UB} &= \frac{1}{M}\left\{\frac{S_2}{k-1} - \frac{S_1}{\nu}\right\} \end{aligned} \tag{1.6}$$

where

$$S_1 = s_{\ell+1}^2 = \mathbf{y}'\mathbf{P}(\mathbf{I}_{N-r} - \mathbf{P}'\mathbf{Z}(\mathbf{Z}'\mathbf{PP}'\mathbf{Z})^{-}\mathbf{Z}'\mathbf{P})\mathbf{P}'\mathbf{y}$$

$$= \mathbf{y}'\mathbf{P_X}\mathbf{y} - \mathbf{y}'\mathbf{P_X}\mathbf{Z}(\mathbf{Z}'\mathbf{P_X}\mathbf{Z})^{-}\mathbf{Z}'\mathbf{P_X}\mathbf{y},$$

$$S_2 = \Sigma_{i=1}^{\ell} s_i^2 = \mathbf{y}'\mathbf{P_X}\mathbf{Z}(\mathbf{Z}'\mathbf{P_X}\mathbf{Z})^{-}\mathbf{Z}'\mathbf{P_X}\mathbf{y}, \tag{1.7}$$

$$M = \Sigma_{i=1}^{\ell} \lambda_i m_i/(k-1) = \mathrm{tr}\mathbf{Z}\mathbf{Z}'\mathbf{P_X}/(k-1)$$

$$= \left\{N - \mathrm{tr}\left[(\mathbf{X}'\mathbf{X})^{-}\Sigma_{i=1}^{k} n_i^2 \overline{\mathbf{x}}_{i\cdot}\overline{\mathbf{x}}_{i\cdot}{}'\right]\right\}/(k-1)$$

for $\mathbf{P_X} = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ and $\overline{\mathbf{x}}_{i\cdot} = \Sigma_{i=1}^{n_i}\mathbf{x}_{ij}/n_i$ (see Seal (1971, pp465-pp467), Battese et al.(1988) and Rao et al.(1993)). Also S_1 and S_2 can be represented as

$$S_1 = \sum_i\sum_j \hat{e}_{ij}^2 \quad \text{and} \quad S_2 = \sum_i\sum_j \hat{u}_{ij}^2 - S_1,$$

where $\{\hat{e}_{ij}\}$ are the residuals from the OLS regression of $y_{ij} - \overline{y}_{i\cdot}$ on $\{x_{ij1} - \overline{x}_{i\cdot 1}, \ldots, x_{ij,p-1} - \overline{x}_{i\cdot p-1}\}$ without the intercept term for $\overline{y}_{i\cdot} = \sum_j y_{ij}/n_i$ and $\overline{x}_{i\cdot\ell} = \sum_j x_{ij\ell}/n_i$, and $\{\hat{u}_{ij}\}$ are the residuals from the OLS regression of y_{ij} on $\{x_{ij1}, \ldots, x_{ij,p-1}\}$ with the intercept term. In the balanced case $n_1 = \cdots = n_k = n$, we have

$$S_1 \sim \sigma_e^2\chi_\nu^2 \tag{1.8}$$

$$S_2 \sim (\sigma_e^2 + n\sigma_v^2)\chi_{k-1}^2$$

and $M = \Sigma_{i=1}^{\ell}\lambda_i m_i/(c-1) = n$. Also in the unbalanced case with $\beta = (\beta_0, 0, \ldots, 0)'$, S_1 and S_2 are also simplified as $S_1 = \Sigma_{i=1}^{k}\Sigma_{j=1}^{n_i}(y_{ij} - \overline{y}_{i\cdot})^2$ and $S_2 = \Sigma_{i=1}^{k} n_i(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2$ for total mean $\overline{y}_{\cdot\cdot}$. So the ANOVA estimators can be much simplified in these special cases. Besides the ANOVA estimators, the minimum norm quadratic unbiased estimators (MINQUE) proposed by Rao(1971a,b) and its modifications are also useful.

These unbiased estimators of σ_v^2, however, have a serious drawback of taking negative values with a positive probability. LaMotte(1973) showed that unbiased nonnegative quadratic estimators of σ_v^2 do not exist. On the other hand, Kleffe and Rao(1986) demonstrated that nonnegative biased quadratic estimators of σ_v^2 fail the minimum condition of consistency as n_i remains fixed, but $k \to \infty$. These prompt us to consider nonnegative estimators other than the quadratic forms.

A simple nonnegative procedure eliminating the undesirable property of an unbiased estimator is a truncation of it at zero, yielding the uniform improvement as noted by Herbach (1959), Thompson (1962) and Berger (1990). The truncated estimator, however, seems still unpleasant because σ_v^2 must be estimated by zero with a probability. Some reasonable procedures have been proposed by Portnoy (1971), Chow and Shao (1988), Mathew et al.(1991), Kubokawa (1994c) and Kubokawa et al.(1993b). In these papers, the mean squared error (MSE) is utilized as a criterion for comparing estimators. However we could not think that the

MSE is an appropriate criterion for evaluating estimators of the scale parameters. What loss function is desirable for estimation of the variance components?

We here discuss the simultanious point estimation of σ_e^2 and σ_v^2 and propose the following loss function for their estimation:

$$L(\hat{\sigma}_e^2, \hat{\sigma}_v^2; \omega) = \nu \left\{ \frac{\hat{\sigma}_e^2}{\sigma_e^2} - \log \frac{\hat{\sigma}_e^2}{\sigma_e^2} - 1 \right\} + (k-1) \left\{ \frac{\hat{\sigma}_e^2 + M\hat{\sigma}_v^2}{\sigma_e^2 + M\sigma_v^2} - \log \frac{\hat{\sigma}_e^2 + M\hat{\sigma}_v^2}{\sigma_e^2 + M\sigma_v^2} - 1 \right\} \tag{1.9}$$

for a couple of unknown parameters $\omega = (\sigma_e^2, \sigma_v^2)$. This can be checked to be a convex function of $\hat{\sigma}_e^2/\sigma_e^2$ and $\hat{\sigma}_v^2/\sigma_v^2$. Note that it incorporates the design matrix and the design parameters. We shall call the loss (1.9) *the Kullback-Leibler Loss* because (1.9) can be derived by the Kulback-Leibler information loss in the balanced case. When a random variable has density $f(x, \theta)$, the Kullback-Leibler distance between θ and the estimator $\hat{\theta}$ is defined by

$$\int \log \left\{ \frac{f(x, \hat{\theta})}{f(x, \theta)} \right\} f(x, \hat{\theta}) dx \tag{1.10}$$

and it is seen that this distance is just $L(\hat{\sigma}_e^2, \hat{\sigma}_v^2; \omega)/2$ in the balanced case of our model.

When we find the best estimators in the sense of minimizing risks within the class of estimators $\hat{\sigma}_e^2 = aS_1$ and $\hat{\sigma}_v^2 = M^{-1}\{bS_2 - aS_1\}$ for constants a and b, they are just the ANOVA estimators $\hat{\sigma}_e^{2UB}$ and $\hat{\sigma}_v^{2UB}$ with respect to the Kullback-Leibler loss (1.9) although the ANOVA estimators are not an optimal choice relative to the MSE criterion.

The main purpose of the present paper is to find nonnegative estimators of (σ_e^2, σ_v^2) improving upon the ANOVA (unbiased) estimators $(\hat{\sigma}_e^{2UB}, \hat{\sigma}_v^{2UB})$ simultaniously relative to the Kullback-Leibler loss (1.9). *The Integral-Expression-of-Risk-Difference*(IERD) method given by Kubokawa(1994a,b) is useful for our purpose. This technique is used in Section 2 to construct a class of estimators better than $(\hat{\sigma}_e^{2UB}, \hat{\sigma}_v^{2UB})$. This class includes various kinds of improved estimators, for example, $(\hat{\sigma}_e^{2EB}, \hat{\sigma}_v^{2EB})$, $(\hat{\sigma}_e^{2PT}, \hat{\sigma}_v^{2PT})$ and $(\hat{\sigma}_e^{2GB}, \hat{\sigma}_v^{2GB})$, which are given by

$$\hat{\sigma}_e^{2EB} = \min \left\{ \frac{S_1}{\nu}, \frac{S_1 + S_2}{\nu + k - 1} \right\}, \tag{1.11}$$

$$\hat{\sigma}_v^{2EB} = \frac{1}{M} \max \left\{ \frac{S_2}{k-1} - \frac{S_1}{\nu}, 0 \right\}, \tag{1.12}$$

$$\hat{\sigma}_e^{2PT} = \min \left\{ \frac{S_1}{\nu}, \frac{S_1 + (k-1)S_2/(k+1)}{\nu + k - 1} \right\}, \tag{1.13}$$

$$\hat{\sigma}_v^{2PT} = \frac{1}{M}\left[\max\left\{\frac{S_2}{k-1}, \frac{S_1+S_2}{\nu+k-3}\right\} - \hat{\sigma}_v^{2PT}\right], \tag{1.14}$$

$$\hat{\sigma}_e^{2GB} = \frac{S_1}{\nu+k-1}\frac{\int_0^{S_2/S_1} x^{(k-1)/2-1}/(1+x)^{(\nu+k-1)/2}dx}{\int_0^{S_2/S_1} x^{(k-1)/2-1}/(1+x)^{(\nu+k+1)/2}dx}, \tag{1.15}$$

$$\hat{\sigma}_v^{2GB} = \frac{1}{M}\left[\frac{S_2}{\nu+k-1}\frac{\int_{S_1/S_2}^{\infty} x^{\nu/2-1}/(1+x)^{(\nu+k-1)/2}dx}{\int_{S_1/S_2}^{\infty} x^{\nu/2-1}/(1+x)^{(\nu+k+1)/2}dx} - \hat{\sigma}_e^{2GB}\right]. \tag{1.16}$$

It is interesting to note that the estimators $(\hat{\sigma}_e^{2EB}, \hat{\sigma}_v^{2EB})$ and $(\hat{\sigma}_e^{2GB}, \hat{\sigma}_v^{2GB})$ can be derived as the empirical Bayes and the generalized Bayes estimators, respectively, in the balanced case. The estimators $(\hat{\sigma}_e^{2PT}, \hat{\sigma}_v^{2PT})$ are always positive, simple and improved procedures. The simulation results given in Section 2.3 demonstrate that $(\hat{\sigma}_e^{2EB}, \hat{\sigma}_v^{2EB})$ and $(\hat{\sigma}_e^{2PT}, \hat{\sigma}_v^{2PT})$ have much smaller risks than $(\hat{\sigma}_e^{2UB}, \hat{\sigma}_v^{2UB})$ and that $(\hat{\sigma}_e^{2GB}, \hat{\sigma}_v^{2GB})$ is the best of the four for $\sigma_v^2/\sigma_e^2 \geq 0.5$ while it has the same risk as $(\hat{\sigma}_e^{2UB}, \hat{\sigma}_v^{2UB})$ at $\sigma_v^2 = 0.0$.

In Section 3, the results of point estimation will be applied to the two-stage GLS estimation of β, to the two-stage GLS F test discussed by Rao et al.(1993), and to the two-stage prediction of a linear combination of β_ℓ's and v_i's, which is related to the small-area estimation. In Section 4, the confidence intervals for σ_e^2, $\sigma_e^2 + n\sigma_v^2$ and σ_v^2 are discussed in the balanced case, and several procedures and their comparison by simulation studies are given. In particular it is clarified that the Kullback-Leibler loss is related to the confidence interval such that the ratio of its end points is minimized.

2. Simultanious Point Estimation

2.1. Derivation of Improved Estimators

We first construct the classes of estimators improving the unbiased ANOVA estimators $(\hat{\sigma}_e^{2UB}, \hat{\sigma}_v^{2UB})$. For this, consider the estimators of the forms

$$\hat{\sigma}_e^2(\psi) = S_1\psi\left(\frac{S_2}{S_1}\right) \tag{2.1}$$
$$\hat{\sigma}_v^2(\phi,\psi) = \frac{1}{M}\left\{S_2\phi\left(\frac{S_1}{S_2}\right) - S_1\psi\left(\frac{S_2}{S_1}\right)\right\}.$$

Then the risk function of $(\hat{\sigma}_e^2(\psi), \hat{\sigma}_v^2(\phi,\psi))$ relative to the Kullback-Leibler loss (1.9) may be written by

$$R(\omega; \hat{\sigma}_e^2(\psi), \hat{\sigma}_v^2(\phi,\psi)) = \nu R_1\left(\omega; S_1\psi\left(\frac{S_2}{S_1}\right)\right) + (k-1)R_2\left(\omega; S_2\phi\left(\frac{S_1}{S_2}\right)\right), \tag{2.2}$$

where

$$R_1(\omega; S_1\psi\left(\frac{S_2}{S_1}\right)) = E\left[\frac{S_1}{\sigma_e^2}\psi\left(\frac{S_2}{S_1}\right) - \log\frac{S_1}{\sigma_e^2}\psi\left(\frac{S_2}{S_1}\right) - 1\right],$$

$$R_2(\omega; S_2\phi\left(\frac{S_1}{S_2}\right)) = E\left[\frac{S_2}{\sigma_e^2 + M\sigma_v^2}\phi\left(\frac{S_1}{S_2}\right) - \log\frac{S_2}{\sigma_e^2 + M\sigma_v^2}\phi\left(\frac{S_1}{S_2}\right) - 1\right].$$

The IERD method can be applied to obtain sufficient conditions for the domination, and we get

Theorem 1. *Assume that*

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

(b) $\psi(w) \geq \psi_0(w)$ *where*

$$\psi_0(w) = \frac{1}{\nu + k - 1}\frac{\int_0^w x^{(k-1)/2-1}/(1+x)^{(\nu+k-1)/2}dx}{\int_0^w x^{(k-1)/2-1}/(1+x)^{(\nu+k+1)/2}dx}. \tag{2.3}$$

Then $R_1(\omega; S_1\psi(S_2/S_1)) \leq R_1(\omega; \nu^{-1}S_1)$ *uniformly for every* ω.

Theorem 2. *Assume that*

(a) $\phi(w)$ *is nondecreasing and* $\phi(0) = (k-1)^{-1}$,

(b) $\phi(w) \leq \phi_0(w)$ *where*

$$\phi_0(w) = \frac{1}{\nu + k - 1}\frac{\int_w^\infty x^{\nu/2-1}/(1+x)^{(\nu+k-1)/2}dx}{\int_w^\infty x^{\nu/2-1}/(1+x)^{(\nu+k+1)/2}dx}. \tag{2.4}$$

Then $R_2(\omega; S_2\phi(S_1/S_2)) \leq R_2(\omega; (k-1)^{-1}S_2)$ *uniformly for every* ω.

The proofs are so technical and given in the Appendix. Combining Theorems 1 and 2 gives that the estimators $(\hat{\sigma}_e^2(\psi), \hat{\sigma}_v^2(\phi,\psi))$ are better than $(\hat{\sigma}_e^{2UB}, \hat{\sigma}_v^{2UB})$ if ϕ and ψ satisfy the conditions (a) and (b) of both theorems.

For (2.3), we can show that

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

for $\alpha > 0$ and $\beta > 0$. Putting

$$\psi_1(w) = \min\left\{\nu^{-1}, \frac{1+w}{\nu+k-1}\right\}, \tag{2.5}$$

$$\psi_2(w) = \min\left\{\nu^{-1}, \frac{1+(k-1)w/(k+1)}{\nu+k-1}\right\},$$

we see that $\psi_1(w)$, $\psi_2(w)$ satisfy the conditions (a) and (b) of Theorem 1, so that for σ_e^2, we get the simple and improved estimators $\hat{\sigma}_e^{2EB}$, $\hat{\sigma}_e^{2PT}$ given by (1.11), (1.13). It is also checked that $\psi_0(w)$ satisfies (a), which yields the smooth improved estimator $\hat{\sigma}_e^{2GB} = \hat{\sigma}_e^2(\psi_0) = S_1\psi_0(S_2/S_1)$ given by (1.15).

For Theorem 2, on the other hand, it can be shown that

$$\frac{\int_w^\infty x^\alpha/(1+x)^\beta dx}{\int_w^\infty x^\alpha/(1+x)^{\beta+1}dx} \geq \frac{\int_w^\infty 1/(1+x)^\beta dx}{\int_w^\infty 1/(1+x)^{\beta+1}dx} = \frac{\beta}{\beta-1}(1+w) \geq 1+w$$

for $\alpha > 0$ and $\beta - \alpha > 1$. Putting

$$\phi_1(w) = \max\left\{\frac{1}{k-1}, \frac{1+w}{\nu+k-1}\right\}, \quad (2.6)$$
$$\phi_2(w) = \max\left\{\frac{1}{k-1}, \frac{1+w}{\nu+k-3}\right\},$$

we can verify that $\phi_1(w)$, $\phi_2(w)$ satisfy the conditions (a) and (b) of Theorem 2. Also these conditions hold for the smooth function $\phi_0(w)$. Combining these results and improved estimators of σ_e^2 yields the superior estimators $\hat{\sigma}_v^{2EB}$, $\hat{\sigma}_v^{2PT}$ and $\hat{\sigma}_v^{2GB}$ given by (1.12), (1.14) and (1.16), respectively. It should be noted that $\hat{\sigma}_v^{2EB}$ is nonnegative, $\hat{\sigma}_v^{2PT}$ is positive and $\hat{\sigma}_v^{2GB}$ is positive and smooth. $\hat{\sigma}_v^{2GB}$ may be complicated to compute it because of including the ratio of integrals or infinite series. $\hat{\sigma}_v^{2EB}$ is a usual truncated procedure. $\hat{\sigma}_v^{2PT}$ is a positive, simple and improved estimator.

2.2. Some Bayesian Properties

We here treat the balanced case $n_1 = \cdots = n_k = n$ and show that $(\hat{\sigma}_e^{2GB}, \hat{\sigma}_v^{2GB})$ and $(\hat{\sigma}_e^{2EB}, \hat{\sigma}_v^{2EB})$ are the generalized Bayes and the empirical Bayes estimators, respectively, relative to the Kullback-Leibler loss.

Let $\eta = 1/\sigma_e^2$, and $\xi = \sigma_e^2/(\sigma_e^2 + n\sigma_v^2)$. From the loss (1.9), the Bayes estimators are generally given by

$$\hat{\sigma}_e^{2B} = \frac{1}{E[\eta|S_1, S_2]},$$
$$\hat{\sigma}_v^{2B} = \frac{1}{n}\left\{\frac{1}{E[\xi\eta|S_1, S_2]} - \hat{\sigma}_e^{2B}\right\}$$

where $E[\cdot|S_1, S_2]$ designates a posterior expectation given S_1 and S_2. For the generalized Bayesness of $(\hat{\sigma}_e^{2GB}, \hat{\sigma}_v^{2GB})$, assume the improper prior distribution

$\eta^{-1}\xi^{-1}d\eta d\xi$, $0 < \eta < \infty$, $0 < \xi < 1$. Then the posterior density of (η, ξ) given S_1 and S_2 is proportional to

$$\xi^{(k-1)/2-1}\eta^{(\nu+k-1)/2-1}e^{-\frac{1}{2}(S_1+\xi S_2)\eta},$$

so that it can be easily checked that $\hat{\sigma}_e^{2GB}$ and $\hat{\sigma}_v^{2GB}$ are derived as the generalized Bayes rules.

For the empirical Bayesness of $(\hat{\sigma}_e^{2EB}, \hat{\sigma}_v^{2EB})$, assume the improper prior distribution $\eta^{-1}d\eta$ with unknown fixed ξ, $0 < \xi < 1$. Then the posterior density of η given S_1 and S_2, and the marginal density of S_1 and S_2 are given by

$$\text{(posterior density)} \propto \eta^{(\nu+k-1)/2-1}e^{-\frac{1}{2}(S_1+\xi S_2)\eta}$$
$$\text{(marginal density)} \propto \xi^{(k-1)/2}[S_1 + \xi S_2]^{-(\nu+k-1)/2}S_1^{\nu/2-1}S_2^{(k-1)/2-1}.$$

Hence the Bayes estimators of σ_e^2 and σ_v^2 are

$$\hat{\sigma}_e^{2B}(\xi) = \frac{S_1 + \xi S_2}{\nu + k - 1},$$
$$\hat{\sigma}_v^{2B}(\xi) = \frac{1}{n}\left\{\frac{S_1 + \xi S_2}{(\nu + k - 1)\xi} - \hat{\sigma}_e^{2B}(\xi)\right\}.$$

Since $0 < \xi < 1$ is unknown, it should be estimated from the marginal density. The maximum likelihood estimator of ξ is written by $\hat{\xi}^{ML} = \min\{(k-1)S_1/(\nu S_2),\ 1\}$, which is substituted in the above Bayes estimators so as to obtain the empirical Bayes rules $\hat{\sigma}_e^{2B}(\hat{\xi}^{ML})$ and $\hat{\sigma}_v^{2B}(\hat{\xi}^{ML})$, just being $\hat{\sigma}_e^{2EB}$ and $\hat{\sigma}_v^{2EB}$ given by (1.11) and (1.12). In this way, the interesting Bayesian interpretations for some estimators are presented.

2.3. Simulation Study

It is of interest to compare, in terms of risk, the estimators $(\hat{\sigma}_e^{2UB}, \hat{\sigma}_v^{2UB})$, $(\hat{\sigma}_e^{2EB}, \hat{\sigma}_v^{2EB})$, $(\hat{\sigma}_e^{2PT}, \hat{\sigma}_v^{2PT})$ and $(\hat{\sigma}_e^{2GB}, \hat{\sigma}_v^{2GB})$, which are here represented as UB, EB, PT and GB estimators, respectively. We treat the following two simple models and provide the results of Monte Carlo simulation for their risk functions relative to the Kullback-Leibler loss.

We first consider the model (1.1) with $\beta = (\beta_0, 0, \ldots, 0)$ and $\sigma_e^2 = 1.0$, that is, $y_{ij} = \beta_0 + v_i + e_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$. Tables 1 and 2 report the average values of their risks based on 50,000 replications in the balanced and the unbalanced cases. It is noticed that the risk of EB estimator has the performance similar to that of PT estimator. They have significant improvements for σ_v^2 near zero. GB estimator may be the best of the four for $\sigma_v^2 > 0.5$ although it has the same risk as UB estimator at $\sigma_v^2 = 0$.

Another model we deal with is (1.1) for $p = 2$ and $n_1 = \cdots = n_k = n$, that is,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}, \ i = 1, \ldots, k, \ j = 1, \ldots, n \tag{2.7}$$

where $\sigma_e^2 = 1.0$ and $\{x_{ij}\}$ are generated from $\mathcal{N}(10, \sigma_x^2)$ for $\sigma_x = 5.0$. Table 3 reports the average values of the risks based on 50,000 replications for $(n, k) = (10, 3), (2, 10), (5, 10)$, namely, $(\nu, k-1) = (26, 2), (9, 9), (79, 19)$. It is revealed that the risk performances of the estimators are similar to Tables 1 and 2. Since PT and GB estimators are always positive and they have good risk performances in large ranges of σ_v^2, they may be employed for a practical use.

Table 1. Expected Kullback-Leibler losses for UB, EB, PT and GB estimators in the balanced cases

σ_A^2		0.0	0.01	0.05	0.1	0.5	1.0	4.0	9.0
	UB	2.20	2.20	2.20	2.20	2.20	2.20	2.20	2.20
$n = 3$	EB	1.45	1.45	1.46	1.47	1.67	1.82	2.06	2.12
$k = 3$	PT	1.53	1.51	1.48	1.45	1.56	1.71	2.00	2.10
	GB	2.20	2.16	2.00	1.86	1.53	1.54	1.82	1.97
	UB	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
$n = 3$	EB	1.46	1.46	1.48	1.51	1.81	1.96	2.08	2.08
$k = 6$	PT	1.50	1.49	1.46	1.47	1.73	1.91	2.07	2.08
	GB	2.08	2.04	1.87	1.73	1.55	1.68	2.00	2.07
	UB	2.17	2.17	2.17	2.17	2.17	2.17	2.17	2.17
$n = 6$	EB	1.35	1.36	1.39	1.45	1.76	1.91	2.09	2.13
$k = 3$	PT	1.37	1.37	1.37	1.41	1.71	1.88	2.08	2.13
	GB	2.17	2.06	1.77	1.58	1.45	1.61	1.95	2.06

Table 2. Expected Kullback-Leibler losses for UB, EB, PT and GB estimators in the unbalanced cases

σ_A^2		0.0	0.01	0.05	0.1	0.5	1.0	4.0	9.0
	UB	2.09	2.09	2.09	2.10	2.12	2.13	2.15	2.16
replications	EB	1.42	1.42	1.47	1.55	1.96	2.07	2.15	2.16
(3,3,5,5,7,7)	PT	1.44	1.43	1.44	1.51	1.92	2.06	2.15	2.16
	GB	2.09	1.98	1.73	1.57	1.65	1.88	2.12	2.15
	UB	2.09	2.09	2.10	2.12	2.24	2.32	2.42	2.44
replications	EB	1.43	1.43	1.47	1.56	2.04	2.23	2.41	2.44
(1,1,5,5,9,9)	PT	1.44	1.44	1.45	1.51	2.00	2.21	2.41	2.44
	GB	2.09	2.00	1.75	1.60	1.70	1.99	2.37	2.43
	UB	2.09	2.09	2.11	2.16	2.50	2.71	2.99	3.07
replications	EB	1.43	1.43	1.47	1.56	2.18	2.54	2.98	3.07
(1,1,1,1,13,	PT	1.44	1.44	1.45	1.52	2.13	2.51	2.97	3.07
13)	GB	2.09	2.01	1.81	1.67	1.81	2.20	2.88	3.05

Table 3. Expected Kullback-Leibler losses for UB, EB, PT and GB estimators in the simple regression models for $\sigma_x = 5.0$

σ_A^2		0.0	0.01	0.05	0.1	0.5	1.0	4.0	9.0
	UB	2.17	2.17	2.18	2.18	2.18	2.19	2.19	2.19
$n = 10$	EB	1.36	1.36	1.42	1.52	1.89	2.01	2.13	2.17
$k = 3$	PT	1.37	1.36	1.40	1.49	1.86	2.00	2.12	2.17
	GB	2.17	2.01	1.63	1.45	1.56	1.76	2.01	2.13
	UB	2.08	2.08	2.08	2.08	2.09	2.10	2.10	2.10
$n = 2$	EB	1.54	1.54	1.55	1.58	1.82	1.98	2.09	2.10
$k = 10$	PT	1.60	1.59	1.56	1.55	1.71	1.91	2.08	2.10
	GB	2.08	2.04	1.94	1.85	1.64	1.71	1.96	2.09
	UB	2.02	2.03	2.03	2.03	2.03	2.03	2.04	2.04
$n = 5$	EB	1.45	1.46	1.59	1.76	2.03	2.03	2.04	2.04
$k = 20$	PT	1.46	1.46	1.56	1.73	2.03	2.03	2.04	2.04
	GB	2.02	1.89	1.60	1.53	1.98	2.03	2.04	2.04

3. Applications

Now we consider utilizing estimators of variance components in applications to the inference of the regression coefficients β.

3.1. Generalized Least Squares Estimation

The estimation of the regression coefficients β is first considered. In the case of known ρ, as stated in Section 1, the GLS estimator of β is given by

$$\hat{\beta}(\rho) = (\mathbf{X}'\mathbf{V}(\rho)^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}(\rho)^{-1}\mathbf{y}$$

where $\mathbf{X}$ is supposed to be with a full rank in this section.

Following Fuller and Battese (1973), we transform the model (1.1) so as to apply the standard OLS methods. Let $\alpha_i = 1 - [(1-\rho)/\{1+(n_i-1)\rho\}]^{1/2}$, $y_{ij}^* = y_{ij} - \alpha_i \overline{y}_{i\cdot}$ and $\mathbf{x}_{ij}^* = \mathbf{x}_{ij} - \alpha_i \overline{\mathbf{x}}_{i\cdot}$, where $\overline{y}_{i\cdot} = \Sigma_j y_{ij}/n_i$ and $\overline{\mathbf{x}}_{i\cdot} = \Sigma_j \mathbf{x}_{ij}/n_i$. The transformed model is written as

$$y_{ij}^* = \mathbf{x}_{ij}^{*\prime}\beta + u_{ij}^*, \quad j = 1,\ldots,n_i,\ i = 1,\ldots,k,$$

or in matrix notation as

$$\mathbf{y}^* = \mathbf{X}^*\beta + \mathbf{u}^*$$

where $\mathbf{u}^* \sim \mathcal{N}_N(\mathbf{0}, \sigma_e^2\mathbf{I}_N)$. Note that $\mathbf{y}^*$ and $\mathbf{X}^*$ depend on ρ. Then the GLS estimator $\hat{\beta}(\rho)$ is represented by $\hat{\beta}(\rho) = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$.

In practice, ρ is rarely known and must be estimated. When ρ is estimated by a consistent estimator $\hat{\rho}$, Fuller and Battese (1973) demonstrated that $\hat{\beta}(\hat{\rho})$ is asymptotically first-order efficient. Based on the estimators of variance components given in the previous sections, we shall consider the following consistent estimators:

$$\hat{\rho}^{EB} = \frac{\hat{\sigma}_v^{2EB}}{\hat{\sigma}_e^{2EB} + \hat{\sigma}_v^{2EB}} = \max(\hat{\rho}^{UB}, 0),$$

$$\hat{\rho}^{PT} = \frac{\hat{\sigma}_v^{2PT}}{\hat{\sigma}_e^{2PT} + \hat{\sigma}_v^{2PT}}, \quad \hat{\rho}^{GB} = \frac{\hat{\sigma}_v^{2GB}}{\hat{\sigma}_e^{2GB} + \hat{\sigma}_v^{2GB}}.$$

It is noted that $\hat{\rho}^{PT}$ and $\hat{\rho}^{GB}$ are always positive and less than 1. Then we investigate the MSE performances of the two-stage GLS estimators $\hat{\beta}(\hat{\rho}^{EB})$, $\hat{\beta}(\hat{\rho}^{PT})$, $\hat{\beta}(\hat{\rho}^{GB})$ and the OLS estimator $\hat{\beta}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}(= \hat{\beta}(0))$.

Table 4 reports the simulation results based on 50,000 replications in the simple regression model (2.7). $\hat{\beta}(\hat{\rho}^{EB})$, $\hat{\beta}(\hat{\rho}^{PT})$ and $\hat{\beta}(\hat{\rho}^{GB})$ have almost same performances of risks, which are stable in comparison with $\hat{\beta}^{OLS}$ as expected. $\hat{\beta}^{OLS}$ has an impermissible large risk for large σ_A^2, especially for $\sigma_x = 1.0$, while it is the best for small σ_A^2.

Table 4. Risks of two-stage GLS estimators $\hat{\beta}(\hat{\rho}^{EB})$, $\hat{\beta}(\hat{\rho}^{PT})$, $\hat{\beta}(\hat{\rho}^{GB})$ and the OLS estimator $\hat{\beta}^{OLS}$ for β in the simple regression models

σ_A^2		0.0	0.01	0.05	0.1	0.5	1.0	4.0	9.0
	$\hat{\beta}(\hat{\rho}^{EB})$	0.19	0.19	0.21	0.23	0.36	0.53	1.20	3.21
$n = 10$	$\hat{\beta}(\hat{\rho}^{PT})$	0.19	0.19	0.21	0.23	0.36	0.53	1.20	3.21
$k = 3$	$\hat{\beta}(\hat{\rho}^{GB})$	0.19	0.19	0.21	0.23	0.36	0.53	1.20	3.21
$\sigma_x = 5.0$	$\hat{\beta}^{OLS}$	0.18	0.18	0.21	0.23	0.40	0.62	1.50	4.12
	$\hat{\beta}(\hat{\rho}^{EB})$	0.31	0.31	0.33	0.34	0.44	0.54	0.81	1.45
$n = 2$	$\hat{\beta}(\hat{\rho}^{PT})$	0.31	0.32	0.33	0.34	0.44	0.54	0.81	1.45
$k = 10$	$\hat{\beta}(\hat{\rho}^{GB})$	0.32	0.33	0.34	0.35	0.44	0.53	0.80	1.45
$\sigma_x = 5.0$	$\hat{\beta}^{OLS}$	0.29	0.30	0.31	0.33	0.46	0.62	1.27	3.22
	$\hat{\beta}(\hat{\rho}^{EB})$	0.05	0.05	0.06	0.06	0.08	0.11	0.21	0.52
$n = 5$	$\hat{\beta}(\hat{\rho}^{PT})$	0.05	0.05	0.06	0.06	0.08	0.11	0.21	0.52
$k = 20$	$\hat{\beta}(\hat{\rho}^{GB})$	0.05	0.05	0.06	0.06	0.08	0.11	0.21	0.52
$\sigma_x = 5.0$	$\hat{\beta}^{OLS}$	0.05	0.05	0.06	0.06	0.10	0.14	0.32	0.87
	$\hat{\beta}(\hat{\rho}^{EB})$	1.06	1.07	1.11	1.14	1.26	1.31	1.44	1.76
$n = 5$	$\hat{\beta}(\hat{\rho}^{PT})$	1.06	1.07	1.11	1.14	1.26	1.31	1.44	1.76
$k = 20$	$\hat{\beta}(\hat{\rho}^{GB})$	1.07	1.08	1.11	1.14	1.26	1.31	1.44	1.76
$\sigma_x = 1.0$	$\hat{\beta}^{OLS}$	1.06	1.07	1.11	1.16	1.58	2.10	4.19	10.46

3.2. Generalized Least Squares F Test

Wu et al.(1988) and Rao et al.(1993) discussed the problem of testing the linear hypothesis $H_0: \mathbf{C}\beta = \mathbf{b}$, where $\mathbf{C}$ is a known $q \times p$ matrix of rank $q(< p)$ and $\mathbf{b}$ is a known $q \times 1$ vector.

Let $\hat{\beta}^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}$ and $\mathbf{X}^*_{\mathbf{C}} = \mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{C}'$. For known ρ, the GLS F test is given by

$$F(\rho) = \frac{(\mathbf{C}\hat{\beta}^* - \mathbf{b})'(\mathbf{X}^{*\prime}_{\mathbf{C}}\mathbf{X}^*_{\mathbf{C}})^{-1}(\mathbf{C}\hat{\beta}^* - \mathbf{b})/q}{(\mathbf{y}^* - \mathbf{X}^*\hat{\beta}^*)'(\mathbf{y}^* - \mathbf{X}^*\hat{\beta}^*)/(N-p)}$$

which has an exact F distribution with q and $N - p$ degrees of freedom.

Since ρ is unknown, two-stage test statistics substituting estimators of ρ are suggested. We investigate the actual type I error rate (size) and the power of the two-stage test statistics $F(\hat{\rho}^{EB})$, $F(\hat{\rho}^{PT})$ and $F(\hat{\rho}^{GB})$ and the OLS F test

$$F^{OLS} = \frac{(\mathbf{C}\hat{\beta}^{OLS} - \mathbf{b})'(\mathbf{X}'_{\mathbf{C}}\mathbf{X}_{\mathbf{C}})^{-1}(\mathbf{C}\hat{\beta}^{OLS} - \mathbf{b})/q}{(\mathbf{y} - \mathbf{X}\hat{\beta}^{OLS})'(\mathbf{y} - \mathbf{X}\hat{\beta}^{OLS})/(N-p)} (= F(0)).$$

for $\mathbf{X}_{\mathbf{C}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$.

For the problem of testing $H_0 : \beta_0 = \beta_1 = 0$ in the simple regression model (2.7), we provide the simulation results for the sizes and the powers of the two-stage GLS F tests and the OLS F test. The expected sizes of the tests for the nominal 5% level based on 50,000 replications are reported in Table 5, and the following observations are revealed: (1) The two-stage GLS F tests $F(\hat{\rho}^{EB})$, $F(\hat{\rho}^{PT})$ and $F(\hat{\rho}^{GB})$ have similar size-performances, which are stable especially for k more than 10. (2) The sizes of F^{OLS} is much inflated for σ^2_A far from zero. (3) The two-stage GLS F tests have better size performances than F^{OLS} while they are also inflated for large σ^2_A. Comparing the results of two cases $\sigma_x = 5.0$ and $\sigma_x = 1.0$ for $n = 5$, $k = 20$, we see that the inflation of the two-stage GLS F-tests is affected by the singularity of the design matrix $\mathbf{X}$. The simulation results for the powers of the tests are given by Table 6 in the case of $n = 5$, $k = 20$ and $\sigma_x = 5.0$. From these investigations by the simulation, we think that the two-stage GLS F tests may be employed for a practical use instead of F^{OLS} with the remark that their size-performances are not very good for small k or small σ^2_x.

Table 5. Sizes (%) of two-stage GLS F-tests $F(\hat{\rho}^{EB})$, $F(\hat{\rho}^{PT})$, $F(\hat{\rho}^{GB})$ and the OLS F test F^{OLS} for $H_0 : \beta_0 = \beta_1 = 0$, nominal 5% level in the simple regression models

σ_A^2		0.0	0.01	0.05	0.1	0.5	1.0	4.0	9.0
	$F(\hat{\rho}^{EB})$	4.6	5.0	7.0	9.2	17.9	21.1	24.5	25.9
$n = 10$	$F(\hat{\rho}^{PT})$	4.4	4.9	6.8	8.9	17.7	21.1	24.5	25.9
$k = 3$	$F(\hat{\rho}^{GB})$	3.1	3.4	4.5	6.1	14.8	19.2	23.8	25.7
$\sigma_x = 5.0$	F^{OLS}	4.9	5.6	8.2	11.7	28.9	38.4	50.0	55.8
	$F(\hat{\rho}^{EB})$	5.1	5.2	5.4	5.7	6.8	7.3	7.5	7.4
$n = 2$	$F(\hat{\rho}^{PT})$	4.8	4.8	5.1	5.4	6.5	7.2	7.5	7.4
$k = 10$	$F(\hat{\rho}^{GB})$	4.5	4.5	4.7	5.0	5.9	6.6	7.3	7.4
$\sigma_x = 5.0$	F^{OLS}	4.9	5.0	5.3	5.6	7.5	9.2	11.8	13.6
	$F(\hat{\rho}^{EB})$	4.8	5.0	5.5	5.9	6.8	7.0	7.1	7.1
$n = 5$	$F(\hat{\rho}^{PT})$	4.7	4.8	5.4	5.9	6.8	7.0	7.1	7.1
$k = 20$	$F(\hat{\rho}^{GB})$	4.0	4.1	4.6	5.2	6.7	7.0	7.1	7.1
$\sigma_x = 5.0$	F^{OLS}	5.2	5.5	6.7	8.1	16.8	22.2	29.3	32.9
	$F(\hat{\rho}^{EB})$	5.1	5.3	6.3	7.4	13.1	15.8	18.5	19.6
$n = 5$	$F(\hat{\rho}^{PT})$	5.1	5.3	6.3	7.4	13.1	15.8	18.5	19.6
$k = 20$	$F(\hat{\rho}^{GB})$	4.9	5.0	6.1	7.2	13.1	15.8	18.5	19.6
$\sigma_x = 1.0$	F^{OLS}	5.2	5.5	6.7	8.1	16.8	22.2	29.3	32.9

Table 6. Powers (%) of two-stage GLS F-tests $F(\hat{\rho}^{EB})$, $F(\hat{\rho}^{PT})$, $F(\hat{\rho}^{GB})$ and the OLS F test F^{OLS} for $H_0 : \beta_0 = \beta_1 = 0$ in the simple regression model with $n = 5$, $k = 20$ and $\sigma_x = 5.0$

$\beta_0 = \beta_1$		0.00	0.005	0.01	0.03	0.06	0.1
	$F(\hat{\rho}^{EB})$	4.8	7.2	15.7	88.3	100.0	100.0
$\sigma_A^2 = 0.0$	$F(\hat{\rho}^{PT})$	4.7	7.1	15.3	88.0	100.0	100.0
	$F(\hat{\rho}^{GB})$	4.0	6.0	13.0	85.4	100.0	100.0
	F^{OLS}	5.2	7.9	17.1	89.9	100.0	100.0
	$F(\hat{\rho}^{EB})$	5.5	8.0	15.5	83.2	100.0	100.0
$\sigma_A^2 = 0.05$	$F(\hat{\rho}^{PT})$	5.4	7.9	15.3	83.1	100.0	100.0
	$F(\hat{\rho}^{GB})$	4.6	6.7	13.3	80.7	100.0	100.0
	F^{OLS}	6.7	9.6	18.5	86.9	100.0	100.0
	$F(\hat{\rho}^{EB})$	6.8	8.2	12.0	53.1	98.5	100.0
$\sigma_A^2 = 0.5$	$F(\hat{\rho}^{PT})$	6.8	8.2	12.0	53.1	98.5	100.0
	$F(\hat{\rho}^{GB})$	6.8	8.1	11.9	53.0	98.5	100.0
	$F(\hat{\rho}^{EB})$	7.0	7.8	10.7	41.5	94.0	100.0
$\sigma_A^2 = 1.0$	$F(\hat{\rho}^{PT})$	7.0	7.8	10.7	41.5	94.0	100.0
	$F(\hat{\rho}^{GB})$	7.0	7.8	10.7	41.5	94.0	100.0

3.3. Two-Stage Prediction

Now we treat the problem of predicting

$$\tau_i = \overline{\mathbf{x}}_{i\cdot}{}'\beta + v_i, \quad i = 1, \ldots, k.$$

This issue is discussed in Peixoto and Harville (1986) and is related to the small-area estimation which has received considerable attension in recent years. For the details, see Battese et al.(1988), Prasad and Rao(1990) and Ghosh and Rao(1994).

In the case of known ρ, the best linear unbiased predictor of τ_i is given by

$$\hat{\tau}_i(\rho) = \overline{\mathbf{x}}_{i\cdot}{}'\hat{\beta}(\rho) + \frac{n_i\rho}{1-\rho+n_i\rho}(\overline{y}_{i\cdot} - \overline{\mathbf{x}}_{i\cdot}{}'\hat{\beta}(\rho)).$$

Since ρ is unknown, two-stage predictors are considered. We look into the MSE performances of the two-stage predictors $\hat{\tau}_i(\hat{\rho}^{EB})$, $\hat{\tau}_i(\hat{\rho}^{PT})$, $\hat{\tau}_i(\hat{\rho}^{GB})$ and the OLS predictor $\hat{\tau}_i(0)$.

The simulation experiments of the MSE's are done for the simple regression model (2.7) and their results are reported in Table 7. It is revealed that the two-stage predictors have common risk performances and the risk of $\hat{\tau}_1(0)$ gets larger for larger σ_A^2 while it is the smallest at $\sigma_A^2 = 0.0$. These properties are quite the same as in the estimation of β.

Table 7. Risks of two-stage GLS predictors $\hat{\tau}_1(\hat{\rho}^{EB})$, $\hat{\tau}_1(\hat{\rho}^{PT})$, $\hat{\tau}_1(\hat{\rho}^{GB})$, and the OLS predictor $\hat{\tau}_1(0)$ for τ_1 in the simple regression models with $\sigma_x = 5.0$

σ_A^2		0.0	0.01	0.05	0.1	0.5	1.0	4.0	9.0
	$\hat{\tau}_1(\hat{\rho}^{EB})$	0.05	0.06	0.07	0.08	0.10	0.10	0.10	0.10
$n = 10$	$\hat{\tau}_1(\hat{\rho}^{PT})$	0.05	0.06	0.07	0.08	0.10	0.10	0.10	0.10
$k = 3$	$\hat{\tau}_1(\hat{\rho}^{GB})$	0.06	0.07	0.07	0.08	0.09	0.10	0.10	0.10
	$\hat{\tau}_1(0)$	0.04	0.04	0.07	0.10	0.34	0.65	1.88	5.58
	$\hat{\tau}_1(\hat{\rho}^{EB})$	0.13	0.14	0.16	0.20	0.34	0.42	0.47	0.49
$n = 2$	$\hat{\tau}_1(\hat{\rho}^{PT})$	0.14	0.15	0.17	0.19	0.33	0.41	0.47	0.49
$k = 10$	$\hat{\tau}_1(\hat{\rho}^{GB})$	0.18	0.19	0.20	0.22	0.32	0.38	0.46	0.49
	$\hat{\tau}_1(0)$	0.07	0.08	0.11	0.16	0.49	0.90	2.56	7.52
	$\hat{\tau}_1(\hat{\rho}^{EB})$	0.02	0.03	0.06	0.08	0.15	0.17	0.19	0.20
$n = 5$	$\hat{\tau}_1(\hat{\rho}^{PT})$	0.02	0.03	0.06	0.08	0.15	0.17	0.19	0.20
$k = 20$	$\hat{\tau}_1(\hat{\rho}^{GB})$	0.03	0.04	0.06	0.08	0.15	0.17	0.19	0.20
	$\hat{\tau}_1(0)$	0.01	0.02	0.06	0.10	0.47	0.94	2.78	8.33

4. Confidence Intervals

Various methods of constructing confidence intervals for the variance components and the ratio have been presented in the literature. For the ratio of variances, exact confidence regions were obtained by Hartley and Rao(1967) and Broemeling(1969). For the 'between' component σ_v^2, Healy(1961) established the exact confidence interval by utilizing an artificial randomization device in addition to the experimental data. Although an exact confidence interval only based on S_1 and S_2 may be desirable, the problem is said to be difficult to settle because the confidence interval of σ_v^2 are truncated at zero and it is not easy to obtain the exact coverage probability. Thereby a variety of approximate intervals was proposed and discussed. Of these, Boardman (1974) investigated and compared their coverage probabilities through simulation experiments and concluded that Moriguti-Bulmer's interval (Moriguti(1954), Bulmer(1957)) and Williams-Tukey's interval (Williams(1962), Tukey(1951)) should be used.

We here treat the balanced case $n_1 = \cdots = n_k = n$ in the model (1.1) where the distributions of S_1 and S_2 are given in (1.8), and we propose some confidence intervals for σ_e^2, $\sigma_e^2 + n\sigma_v^2$ and σ_v^2.

4.1. Confidence Intervals of σ_e^2 and $\sigma_e^2 + n\sigma_v^2$

For the interval estimation of σ_e^2, Tate and Klett(1959) derived three types of intervals: the shortest unbiased confidence interval, the minimum length confidence interval and the equal tailed confidence interval. While the criterion of minimizing the length of a confidence interval is needed for a location parameter, it may not be appropriate for the scale parameter. In the interval estimation of the scale parameter, it may be reasonable to request minimizing the ratio of the end points of a confidence interval. We shall call such a confidence interval *the Minimum Ratio Confidence Interval* (MRCI), which is, in this case, given by

$$I_e^{MR} = \left[\frac{S_1}{\nu a_1}, \frac{S_1}{\nu a_2} \right] \tag{4.1}$$

where a_1 and a_2 ($a_1 > a_2$) satisfy

$$a_1 - a_2 = \log(a_1/a_2), \quad P[\nu a_2 \le \chi_\nu^2 \le \nu a_1] = 1 - \alpha$$

for the confidence coefficient α, $0 < \alpha < 1$. This MRCI I_e^{MR} is identical to the shortest unbiased confidence interval where various values of a_1 and a_2 are given in the tables of Lindley, East and Hamilton(1960). It is interesting to note that I_e^{MR} is also motivated from the Kullback-Leibler loss for the unbiased estimator

$\hat{\sigma}_e^{2UB} = S_1/\nu$, that is, the set of the form

$$\left\{\sigma_e^2;\ \frac{S_1/\nu}{\sigma_e^2} - \log\frac{S_1/\nu}{\sigma_e^2} - 1 \leq a_1 - \log a_1 - 1 = a_2 - \log a_2 - 1,\ a_2 < a_1\right\} \quad (4.2)$$

is identical to I_e^{MR} as easily shown.

For improving on I_e^{MR} by use of S_2, consider the confidence interval of the form

$$I_e(\psi) = \left[\frac{S_1}{a_1}\psi\left(\frac{S_2}{S_1}\right), \frac{S_1}{a_2}\psi\left(\frac{S_2}{S_1}\right)\right] \quad (4.3)$$

which has the same ratio of the end points as I_e^{MR}. Then we can get the following theorem which can be proved by combining the proof of Theorem 1 and Section 3 of Kubokawa (1994a).

Theorem 3. *Assume that*

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

(b) $\psi(w)$ *satisfies the inequality:*

$$f_{\nu+2}\left(\frac{a_1}{\psi(w)}\right)F_{k-1}\left(\frac{a_1 w}{\psi(w)}\right) \geq f_{\nu+2}\left(\frac{a_2}{\psi(w)}\right)F_{k-1}\left(\frac{a_2 w}{\psi(w)}\right) \quad (4.4)$$

where $f_n(x)$ *designates the density of* χ_n^2 *and* $F_n(x) = \int_0^x f_n(t)dt$.
Then $P[\sigma_e^2 \in I_e(\psi)] \geq P[\sigma_e^2 \in I_e^{MR}]$ *uniformly for every* $\omega = (\sigma_e^2, \sigma_v^2)$.

Define $\psi_0(w)$ by a solution of the equation

$$f_{\nu+2}\left(\frac{a_1}{\psi(w)}\right)F_{k-1}\left(\frac{a_1 w}{\psi(w)}\right) = f_{\nu+2}\left(\frac{a_2}{\psi(w)}\right)F_{k-1}\left(\frac{a_2 w}{\psi(w)}\right)$$

and let

$$\psi_1(w) = \min\left\{\nu^{-1}, \frac{1+w}{\nu+k-1}\right\}.$$

The discussions given in Section 3 of Kubokawa (1994a) can show that $\psi_0(w)$ and $\psi_1(w)$ satisfy the conditions (a) and (b) of Theorem 3, and we get the improved confidence intervals $I_e(\psi_0)$ and

$$I_e^{MR*} = I_e(\psi_1) = \left[\frac{1}{a_1}\min\left\{\frac{S_1}{\nu}, \frac{S_1+S_2}{\nu+k-1}\right\}, \frac{1}{a_2}\min\left\{\frac{S_1}{\nu}, \frac{S_1+S_2}{\nu+k-1}\right\},\right].$$

Similar to Section 2.2, it can be shown that $I_e(\psi_0)$ is the generalized Bayes confidence interval among the interval estimators of the form $I_e(\psi)$, (4.3), against the improper prior distribution $\eta^{-1}\xi^{-1}d\eta d\xi$ for $\eta = 1/\sigma^2$ and $\xi = \sigma_e^2/(\sigma_e^2 + n\sigma_v^2)$. That is, $\psi_0(w)$ is a function of maximizing the posterior probability

$$\int_0^1\int_{a_2/S_1\psi}^{a_1/S_1\psi} \xi^{(k-1)/2-1}\eta^{(\nu+k-1)/2-1}e^{-\frac{1}{2}(S_1+\xi S_2)\eta}d\eta d\xi.$$

Similarly, the Bayes solution against the improper prior $\eta^{-1}d\eta$ with unknown fixed ξ is given by

$$\psi^B(w,\xi) = \frac{(a_2 - a_1)(1+\xi w)}{(\nu + k - 1)\log(a_2/a_1)} = \frac{1+\xi w}{\nu + k - 1},$$

so that the MLE of ξ in the marginal distribution, $\hat{\xi}^{ML} = \min\{(k-1)/(\nu w), 1\}$, is substituted to get $\psi^{EB}(w) = \psi^B(w, \hat{\xi}^{ML})$, just being $\psi_1(w)$. In this way, $I_e^{MR*} = I_e(\psi_1)$ is interpreted as an empirical Bayes confidence interval.

The same arguments can be applied to construct a confidence interval for $\sigma_e^2 + n\sigma_v^2$. The MRCI is given by

$$I_{e+v}^{MR} = \left[\frac{S_2}{(k-1)b_1}, \frac{S_2}{(k-1)b_2}\right] \tag{4.5}$$

where $b_1 - b_2 = \log(b_1/b_2)$ and $P[\sigma_e^2 + n\sigma_v^2 \in I_{e+v}^{MR}] = 1 - \alpha$. The following theorem guarantees the existence of better confidence intervals within the class of confidence intervals

$$I_{e+v}(\phi) = \left[\frac{S_2}{b_1}\phi\left(\frac{S_1}{S_2}\right), \frac{S_2}{b_2}\phi\left(\frac{S_1}{S_2}\right)\right]. \tag{4.6}$$

Theorem 4. *Assume that*

(a) $\phi(w)$ is nondecreasing and $\phi(0) = (k-1)^{-1}$,

(b) $\phi(w)$ satisfies the inequality:

$$\left\{1 - F_\nu\left(\frac{b_1 w}{\phi(w)}\right)\right\} f_{k+1}\left(\frac{b_1}{\phi(w)}\right) \le \left\{1 - F_\nu\left(\frac{b_2 w}{\phi(w)}\right)\right\} f_{k+1}\left(\frac{b_2}{\phi(w)}\right). \tag{4.7}$$

Then $P[\sigma_e^2 + n\sigma_v^2 \in I_{e+v}(\phi)] \ge P[\sigma_e^2 + n\sigma_v^2 \in I_{e+v}^{MR}]$ uniformly for every ω.

Let $\phi_0(w)$ be a solution of the equality in (4.7), and let

$$\phi_1(w) = \max\left\{(k-1)^{-1}, \frac{1+w}{\nu + k - 1}\right\}.$$

From Theorem 4, it can be verified that they present improved confidence intervals $I_{e+v}(\phi_0)$ and

$$I_{e+v}^{MR*} = I_{e+v}(\phi_1),$$

and that they are also the generalized Bayes and the empirical Bayes confidence intervals, respectively.

The expected values of coverage probabilities of (I_e^{MR}, I_e^{MR*}) and $(I_{e+v}^{MR}, I_{e+v}^{MR*})$ are given in Tables 8 and 9 based on the simulation experiments with 50,000 replications. I_e^{MR*} and I_{e+v}^{MR*} have highest coverage probabilities at $\sigma_A^2 = 0.0$. In particular, I_{e+v}^{MR*} has a higher gain in coverage probability than I_e^{MR*}.

Table 8. Coverage probabilities (%) of the confidence intervals I_e^{MR} and I_e^{MR*} for the variance component σ_e^2 in the balanced case

σ_A^2		0.0	0.1	0.5	1.0	3.0	9.0
$n=3,k=3$	I_e^{MR}	95.0	95.0	95.1	95.1	95.1	95.1
$\nu=3$	I_e^{MR*}	96.0	96.0	95.8	95.6	95.3	95.1
$n=2,k=6$	I_e^{MR}	95.1	95.1	95.1	95.1	95.1	95.1
$\nu=3$	I_e^{MR*}	96.1	96.1	96.0	95.8	95.3	95.1
$n=3,k=3$	I_e^{MR}	95.1	95.1	95.1	95.1	95.1	95.1
$\nu=6$	I_e^{MR*}	95.9	95.9	95.6	95.5	95.2	95.1

Table 9. Coverage probabilities (%)of the confidence intervals I_{e+v}^{MR} and I_{e+v}^{MR*} for $\sigma_e^2 + n\sigma_v^2$ in the balanced case

σ_A^2		0.0	0.1	0.5	1.0	3.0	9.0
$n=3,k=3$	I_{e+v}^{MR}	95.0	95.0	95.0	95.0	95.0	95.0
$\nu=3$	I_{e+v}^{MR*}	98.6	98.5	98.1	97.5	95.9	95.1
$n=2,k=6$	I_{e+v}^{MR}	95.1	95.1	95.1	95.1	95.1	95.1
$\nu=3$	I_{e+v}^{MR*}	97.6	97.5	96.9	96.3	95.3	95.1
$n=3,k=3$	I_{e+v}^{MR}	94.9	94.9	94.9	94.9	94.9	94.9
$\nu=6$	I_{e+v}^{MR*}	98.8	98.7	98.5	98.1	96.1	94.9

4.2. Confidence Intervals of σ_v^2

As stated in the beginning of Section 4, the simulation study by Boardman (1974) implies that Moriguti-Bulmer's and Williams-Tukey's confidence intervals have desirable coverage probabilities for the confidence coefficient. We here treat the Williams-Tukey type confidence interval of σ_v^2 for the simplicity of the form and derive several other confidence intervals by using the method of Williams (1962).

Let J denote an equal-tails confidence interval of the ratio $(\sigma_e^2 + n\sigma_v^2)/\sigma_e^2$, given by

$$J = \left[\frac{S_2}{c_1 S_1}, \frac{S_2}{c_2 S_1} \right],$$

where $P(\chi_{k-1}^2/\chi_\nu^2 \leq c_2) = P(\chi_{k-1}^2/\chi_\nu^2 \geq c_1) = \alpha/2$. Then the William-Tukey type confidence interval is provided by

$$I_v^{WT} = \left[\frac{(S_2 - c_1 S_1)^+}{n(k-1)b_1}, \frac{(S_2 - c_2 S_1)^+}{n(k-1)b_2} \right], \tag{4.8}$$

where $x^+ = \max(x, 0)$. From Williams(1962) or Figure 1, it is seen that

$$\begin{aligned} P(\sigma_v^2 \in I_v^{WT}) &\geq P(\sigma_e^2 + n\sigma_v^2 \in I_{e+v}^{MR} \text{ and } \frac{\sigma_e^2 + n\sigma_v^2}{\sigma_e^2} \in J) \\ &= P(\sigma_e^2 + n\sigma_v^2 \in I_{e+v}^{MR}) + P(\frac{\sigma_e^2 + n\sigma_v^2}{\sigma_e^2} \in J) \\ &\quad - P(\sigma_e^2 + n\sigma_v^2 \in I_{e+v}^{MR} \text{ or } \frac{\sigma_e^2 + n\sigma_v^2}{\sigma_e^2} \in J) \\ &> 1 - 2\alpha, \end{aligned}$$

so that the confidence coefficient of I_v^{WT} is guaranteed to be more than $1 - 2\alpha$. By employing I_{e+v}^{MR*} instead of I_{e+v}^{MR}, we can propose the confidence interval

$$I_v^{WT*} = I_v(\phi_1) = \left[\frac{\phi_1(\frac{S_1}{S_2})}{nb_1}(S_2 - c_1 S_1)^+, \ \frac{\phi_1(\frac{S_1}{S_2})}{nb_2}(S_2 - c_2 S_1)^+ \right], \tag{4.9}$$

which has the same ratio of the endpoints as I_v^{WT} for $S_2 > c_1 S_1$. From Figure 1, it is noted that the positive region of I_v^{WT*} is larger than that of I_v^{WT}.

Using I_e^{MR} instead of J with the same idea as above, we can get another type of a confidence interval based on (I_{e+v}^{MR}, I_e^{MR}), given by

$$I_v^{AT} = \left[\frac{1}{n} \left\{ \frac{S_2}{(k-1)b_1} - \frac{S_1}{\nu a_2} \right\}^+, \ \frac{1}{n} \left\{ \frac{S_2}{(k-1)b_2} - \frac{S_1}{\nu a_1} \right\}^+ \right]. \tag{4.10}$$

Also the confidence interval based on $(I_{e+v}^{MR*}, I_e^{MR*})$ is given by

$$\begin{aligned} I_v^{AT*} &= I_v^{AT*}(\phi_1, \psi_1) \\ &= \left[\frac{1}{n} \left\{ \frac{\phi_1(\frac{S_1}{S_2})}{b_1} S_2 - \frac{\psi_1(\frac{S_2}{S_1})}{a_2} S_1 \right\}^+, \ \frac{1}{n} \left\{ \frac{\phi_1(\frac{S_1}{S_2})}{b_2} S_2 - \frac{\psi_1(\frac{S_2}{S_1})}{a_1} S_1 \right\}^+ \right], \end{aligned} \tag{4.11}$$

which has a larger positive region than I_v^{AT} as demonstrated in Figure 2.

The simulation results are given in Table 10. It is revealed that I_v^{AT} and I_v^{AT*} have higher coverage probabilities than I_v^{WT} and I_v^{WT*} and that both of I_v^{WT*} and I_v^{AT*} have two peaks in coverage probabilities at $\sigma_A^2 = 0.0$ and near $\sigma_A^2 = 0.5$.

Table 10. Coverage probabilities (%) of the confidence intervals I_v^{WT}, I_v^{WT*}, I_v^{AT} and I_v^{AT*} for the variance component σ_v^2 in the balanced case

σ_A^2		0.0	0.01	0.05	0.1	0.5	1.0	3.0	9.0
	I_v^{WT}	97.5	95.2	95.7	96.0	96.3	96.2	95.8	95.5
$n=3$	I_v^{WT*}	97.5	95.3	96.1	96.9	98.3	98.1	96.6	95.5
$k=3$	I_v^{AT}	99.3	98.6	98.6	98.5	97.5	97.0	96.4	95.9
$\nu=3$	I_v^{AT*}	99.3	98.7	98.9	99.2	99.3	98.9	97.2	95.9
	I_v^{WT}	97.5	95.2	95.5	96.0	96.9	96.9	96.5	96.1
$n=2$	I_v^{WT*}	97.5	95.3	95.8	96.5	98.0	97.8	96.7	96.1
$k=6$	I_v^{AT}	99.0	98.7	98.8	98.9	98.6	98.1	97.2	96.5
$\nu=3$	I_v^{AT*}	99.0	99.0	99.1	99.3	99.6	99.0	97.4	96.5
	I_v^{WT}	97.5	95.2	95.6	95.8	96.0	95.8	95.5	95.2
$n=3$	I_v^{WT*}	97.5	95.3	96.2	96.8	98.3	98.2	96.4	95.2
$k=3$	I_v^{AT}	99.7	98.7	98.5	98.3	97.3	96.8	96.1	95.6
$\nu=6$	I_v^{AT*}	99.7	97.0	97.4	97.8	98.8	98.7	96.8	95.6

APPENDIX

Proof of Theorem 1. Since $\lim_{w\to\infty}\psi(w) = \nu^{-1}$, from the IERD method of Kubokawa (1994a,b), we have

$$\begin{aligned}
&R_1(\omega; S_1\nu^{-1}) - R_1(\omega; S_1\psi\left(\frac{S_2}{S_1}\right)) \\
&= E\left[\left\{\frac{S_1}{\sigma_e^2}\psi\left(\frac{S_2}{S_1}t\right) - \log\frac{S_1}{\sigma_e^2}\psi\left(\frac{S_2}{S_1}t\right) - 1\right\}\Big|_{t=1}^{\infty}\right] \qquad (A.1) \\
&= E\left[\int_1^\infty \frac{d}{dt}\left\{\frac{S_1}{\sigma_e^2}\psi\left(\frac{S_2}{S_1}t\right) - \log\frac{S_1}{\sigma_e^2}\psi\left(\frac{S_2}{S_1}t\right) - 1\right\}dt\right].
\end{aligned}$$

Let $v = S_1/\sigma_e^2$, $u_i = s_i^2/(\sigma_e^2 + \lambda_i\sigma_v^2)$, and $\theta_i = 1 + \lambda_i\sigma_v^2/\sigma_e^2$, and denote the density functions of v and u_i by f and g_i, respectively. Carrying out the differentiation in (A.1) gives

$$\begin{aligned}
&E\left[\int_1^\infty \left\{\frac{S_1}{\sigma_e^2} - \frac{1}{\psi(S_2t/S_1)}\right\}\frac{S_2}{S_1}\psi'\left(\frac{S_2}{S_1}t\right)dt\right] \\
&= \int\cdots\int\int_1^\infty \left\{v - \frac{1}{\psi(\Sigma\theta_i u_i v/t)}\right\}(\Sigma\theta_i u_i/v)\,\psi'(\Sigma\theta_i u_i t/v)\,dt \\
&\qquad f(v)\Pi_i g_i(u_i)dv\Pi_i du_i.
\end{aligned}$$

Making the transformations $(t/v)u_i = w_i$ and $1/t = x$ in order, we observe that the r.h.s. of (A.1) is equal to

$$\int\cdots\int\int_1^\infty \left\{v - \frac{1}{\psi(\Sigma\theta_i w_i)}\right\}(\Sigma\theta_i w_i/t)\,\psi'(\Sigma\theta_i w_i t)$$

$$\left(\frac{v}{t}\right)^{\ell} f(v)\Pi_i g_i\left(vw_i/t\right) dtdv\Pi_i dw_i$$

$$= \int \cdots \int \left\{ v - \frac{1}{\psi(\Sigma\theta_i w_i)} \right\} (\Sigma\theta_i w_i)\, \psi'\left(\Sigma\theta_i w_i\right) \tag{A.2}$$

$$v^{\ell} f(v) \int_0^1 x^{\ell-1}\Pi_i g_i\left(w_i v x\right) dxdv\Pi_i dw_i.$$

Since $\psi'(w) \geq 0$, it is concluded that the r.h.s. of (A.2) is nonnegative if

$$\psi(\Sigma\theta_i w_i) \geq \frac{\int_0^\infty v^{\ell} f(v) \int_0^1 x^{\ell-1}\Pi_i g_i(w_i v x) dxdv}{\int_0^\infty v^{\ell+1} f(v) \int_0^1 x^{\ell-1}\Pi_i g_i(w_i v x) dxdv}. \tag{A.3}$$

Since $\theta_i \geq 1$ and $\psi'(w) \geq 0$, it follows that $\psi(\Sigma\theta_i w_i) \geq \psi(\Sigma w_i)$, which, from (A.3), gives the sufficient condition that $\psi(\Sigma w_i)$ is greater than or equal to the r.h.s. of (A.3). Integrating out the r.h.s. of (A.3) with respect to v yields $\psi_0(\Sigma w_i)$ given by (2.3). Hence the inequality (A.3) is guaranteed by the condition (b) of Theorem 1, which is established.

Proof of Theorem 2. Since $\phi(0) = (k-1)^{-1}$, observe that

$$R_2(\omega; S_2(k-1)^{-1}) - R_2(\omega; S_2\phi\left(\frac{S_1}{S_2}\right)) \tag{A.4}$$

$$= -E\left[\int_0^1 \frac{d}{dt}\left\{ \frac{S_2}{\sigma_e^2 + M\sigma_v^2}\phi\left(\frac{S_1}{S_2}t\right) - \log\frac{S_2}{\sigma_e^2 + M\sigma_v^2}\phi\left(\frac{S_1}{S_2}t\right) - 1 \right\} dt\right]$$

$$= \int \cdots \int \int_0^1 \left\{ \frac{1}{\phi(\frac{vt}{\Sigma\theta_i u_i})} - \frac{\Sigma\theta_i u_i}{1+M\tau} \right\} \frac{v}{\Sigma\theta_i u_i}\phi'\left(\frac{vt}{\Sigma\theta_i u_i}\right) dt$$

$$f(u)\Pi_i g_i(u_i) dv\Pi_i du_i$$

for $\tau = \sigma_v^2/\sigma_e^2$. Making the transformations $(t/\Sigma\theta_i u_i)v = w$ and $w(1/t) = y$ in order, we can rewrite (A.4) as

$$\int \cdots \int \int_0^1 \left\{ \frac{1}{\phi(w)} - \frac{\Sigma\theta_i u_i}{1+M\tau} \right\} \phi'(w) w \Sigma\theta_i u_i/t^2$$

$$f(\Sigma\theta_i u_i w/t)\Pi_i g_i(u_i) dt\Pi_i du_i dw$$

$$= \int \cdots \int \int_w^\infty \left\{ \frac{1}{\phi(w)} - \frac{\Sigma\theta_i u_i}{1+M\tau} \right\} \phi'(w)\Sigma\theta_i u_i \tag{A.5}$$

$$f(\Sigma\theta_i u_i y)\Pi_i g_i(u_i) dy\Pi_i du_i dw$$

so that since $\phi'(w) \geq 0$, the l.h.s. of (A.4) is nonnegative if

$$\phi(w) \leq \frac{\int \cdots \int \int_w^\infty (\Sigma\theta_i u_i) f(\Sigma\theta_i u_i y)\Pi_i g_i(u_i) dy\Pi_i du_i}{\int \cdots \int \int_w^\infty (\Sigma\theta_i u_i)^2/(1+M\tau) f(\Sigma\theta_i u_i y)\Pi_i g_i(u_i) dy\Pi_i du_i}. \tag{A.6}$$

Letting $s = \Sigma_{i=1}^{\ell} u_i$ and $z_i = u_i/s$, we see that

$$s \sim \chi^2_{k-1},$$
$$z_i \sim Beta(m_i/2, \Sigma_{j\neq i} m_j/2)$$

and s and z_i are independent. Let $Q = \Sigma\theta_i z_i$, being independent of s. The r.h.s. of (A.6) can be rewritten as

$$\frac{E^Q[\int\int_w^\infty Qsf(Qsy)g(s)dyds]}{E^Q[\int\int_w^\infty Q^2s^2/(1+M\tau)f(Qsy)g(s)dyds]} \tag{A.7}$$
$$= \frac{E^Q[\int_{Qw}^\infty \int sf(sx)g(s)dsdx]}{E^Q[Q/(1+M\tau)\int_{Qw}^\infty \int s^2 f(sx)g(s)dsdx]},$$

where $g(s)$ is a density of χ^2_{k-1}. Since Q and $\int_{Qw}^\infty \int s^2 f(sx)g(s)dsdx$ are monotone in the opposite directions, we can show the following inequality holds for the denominator of the r.h.s. of (A.7):

$$E^Q\left[\frac{Q}{1+M\tau}\int_{Qw}^\infty \int s^2 f(sx)g(s)dsdx\right]$$
$$\leq E^Q\left[\frac{Q}{1+M\tau}\right] E^Q\left[\int_{Qw}^\infty \int s^2 f(sx)g(s)dsdx\right]. \tag{A.8}$$

Here observe that

$$E^Q\left[\frac{Q}{1+M\tau}\right] = \frac{1}{1+M\tau} + \frac{1}{1+M\tau}\sum_{i=1}^{\ell}\lambda_i\tau E[z_i] \tag{A.9}$$
$$= \frac{1}{1+M\tau} + \frac{1}{1+M\tau}\frac{\Sigma\lambda_i\tau m_i}{k-1} = 1,$$

since $M = \sum\lambda_i m_i/(k-1)$. Combining (A.6), (A.7), (A.8) and (A.9) gives a sufficient condition as

$$\phi(w) \leq \frac{E^Q[\int_{Qw}^\infty \int sf(sx)g(s)dsdx]}{E^Q[\int_{Qw}^\infty \int s^2 f(sx)g(s)dsdx]}. \tag{A.10}$$

Furthermore the r.h.s. of (A.10) can be shown to be greater than or equal to

$$\inf_Q\left\{\frac{\int_{Qw}^\infty \int sf(sx)g(s)dsdx}{\int_{Qw}^\infty \int s^2 f(sx)g(s)dsdx}\right\} = \inf_Q\{\phi_0(Qw)\} \tag{A.11}$$

where the equality can be obtained by integrating out the l.h.s. of (A.11) with respect to s. As noted below Theorem 2, $\phi_0(w)$, given by (2.4), is a nondecreasing function and $Q \geq 1$. Hence $\phi_0(Qw) \geq \phi_0(w)$. Combining this inequality, (A.10) and (A.11), we get the sufficient condition that $\phi(w) \leq \phi_0(w)$, which is just the condition (b), and the proof of Theorem 2 is complete.

Acknowledgement. This reseach has been supported by the NSERC grant no. A3088, CANADA, during the visit of the first and the third authors at Carleton University, Ottawa. I am grateful to Professor J.N.K. Rao for his valuable suggestions, which motivated us to tackle the issues of this paper.

REFERENCES

Battese, G.E., Harter, R.M. and Fuller, W.A.(1988). An error-components model for prediction of county crop areas usin survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28-36.

Berger, J.O.(1990). On the admissibility of unbiased estimators. *Statist. Probab. Letters*, **9**, 381-384.

Boardman, T.J.(1974). Confidence intervals for variance components - A comparative Monte Carlo study. it Biometrics, **30**, 251-262.

Broemeling, L.D.(1969). Confidence regions for variance ratios of random models. *J. Amer. Statist. Assoc.*, **64**, 660-664.

Bulmer, M.G.(1957). Approximate confidence limits for components of variance. *Biometrika*, **44**, 159-167.

Chow, S.C. and Shao, J.(1988). A new procedure for the estimation of variance components. *Statist. Probab. Letters*, **6**, 349-355.

Fuller, W.A. and Battese, G.E.(1973). Transformations for estimation of linear models with nested-error structure. *J. Amer. Statist. Assoc.*, **68**, 626-632.

Ghosh, M. and Rao, J.N.K.(1994). Small area estimation: An appraisal. *Statist. Sci.*, **9**, 55-93.

Hartley, H.O. and Rao, J.N.K.(1967). Maximum likelihood estimation for mixed analysis of variance model. *Biometrika*, **54**, 93-108.

Harville, D.A.(1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.*, **72**, 320-340.

Healy, W.C.(1961). Limits for a variance component with an exact confidence coeficient. *Ann. Math. Statist.*, **32**, 466-476.

Henderson, C.R.(1953). Estimation of variance and covariance components. *Biometrics*, **9**, 226-252.

Herbach, L.H.(1959). Properties of model II-type analysis of variance tests, A: Optimum nature of the F-test for model II in the balanced case. *Ann. Math. Statist.*, **30**, 939-959.

Kleffe, J. and Rao, J.N.K.(1986). The existence of asymptotically unbiased nonnegative quadratic estimates of variance components in ANOVA models. *J. Amer. Statist. Assoc.* **81**, 692-698.

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

Kubokawa,T.(1994b). Double shrinkage estimation of ratio of scale parameters. *Ann. Inst. Statist. Math.*, **46**, 95-116.

Kubokawa, T.(1995). Estimation of variance components in mixed linear models. *J. Multivariate Anal.*, **53**, 210-236.

Kubokawa, T. and Robert, C.P.(1994). New perspectives on linear calibration. *J. Multivariate Anal.*, **51**, 178-200.

Kubokawa, T., Robert, C.P. and Saleh, A.K.Md.E.(1993a). Estimation of noncentrality parameters. *Canad. J. Statist.*, **21**, 45-57.

Kubokawa, T., Saleh, A.K.Md.E. and Makita, S.(1993b). On improved positive estimators of variance components. *Statistics Decisions*, Supplement Issue 3, 1-16.

LaMotte, L.R.(1976). Invariant quadratic estimators in the random, one-way ANOVA model. *Biometrics* **32**, 793-804.

Lindley, D.V., East, D.A. and Hamilton, P.A.(1960). Tables for making inferences about the variance of a normal distribution. *Biometrika* **47**, 433-437.

Mathew, T., Sinha, B.K. and Sutradhar, B.C.(1991). Nonnegative estimation of random effects variance components in balanced mixed models. *Proceedings of the International Symposium on Nonparametric Statistics and Related Topics*(ed. A.K.Md.E. Saleh).

Mathew, T., Sinha, B.K. and Sutradhar, B.C.(1992). Nonnegative estimation of variance components in unbalanced mixed models with two variance components. *J. Multivariate Anal.*, **42**, 77-101.

Moriguti, S.(1954). Confidence Limits for a variance component. *Rep. Statist. Appl. Res. JUSE*, **3**, 7-19.

Peixoto, J.L. and Harville, D.A.(1986). Comparisons of alternative predictors under the balanced one-way random model. *J. Amer. Statist. Assoc.*, **81**, 431-436.

Portnoy, S.(1971). Formal Bayes estimation with application to a random effect model. *Ann. Math. Statist.* **42**, 1379-1402.

Prasad, N.G.N. and Rao, J.N.K.(1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.*, **85**, 163-171.

Rao, C.R.(1971a). Estimation of variance and covariance components-MINQUE theory. *J. Multivariate Anal.*, **1**, 257-275.

Rao, C.R.(1971b). Minimum variance quadratic unbiased estimation of variance components. *J. Multivariate Anal.*, **1**, 445-456.

Rao, J.N.K., Sutradhar, B.C. and Yue, K.(1993). Generalized least squares F test in regression analysis with two-stage cluster samples. *J. Amer. Statist. Assoc.*, **88**, 1388-1391.

Searle, S.R.(1971). *Linear Models*, John Wiley and Sons, New York.

Searle, S.R., Casella, G. and McCulloch, C.E.(1992). *Variance Components.* Wiley, New York.

Tate, R.F. and Klett, G.W.(1959). Optimal confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.*, **54**, 674-682.

Thompson, W.A.Jr.(1962). The problem of negative estimates of variance components. *Ann. Math. Statist.*, **33**, 273-289.

Tukey, J.W.(1951). Components in regression. *Biometrics*, **7**, 33-69.

Williams, J.S.(1962). A confidence interval for variance components. *Biometrika*, **49**, 278-281.

Wu, C.F.J., Holt, D. and Holmes, D.J.(1988). The effect of two-stage sampling on the F statistic. *J. Amer. Statist. Assoc.*, **83**, 150-159.

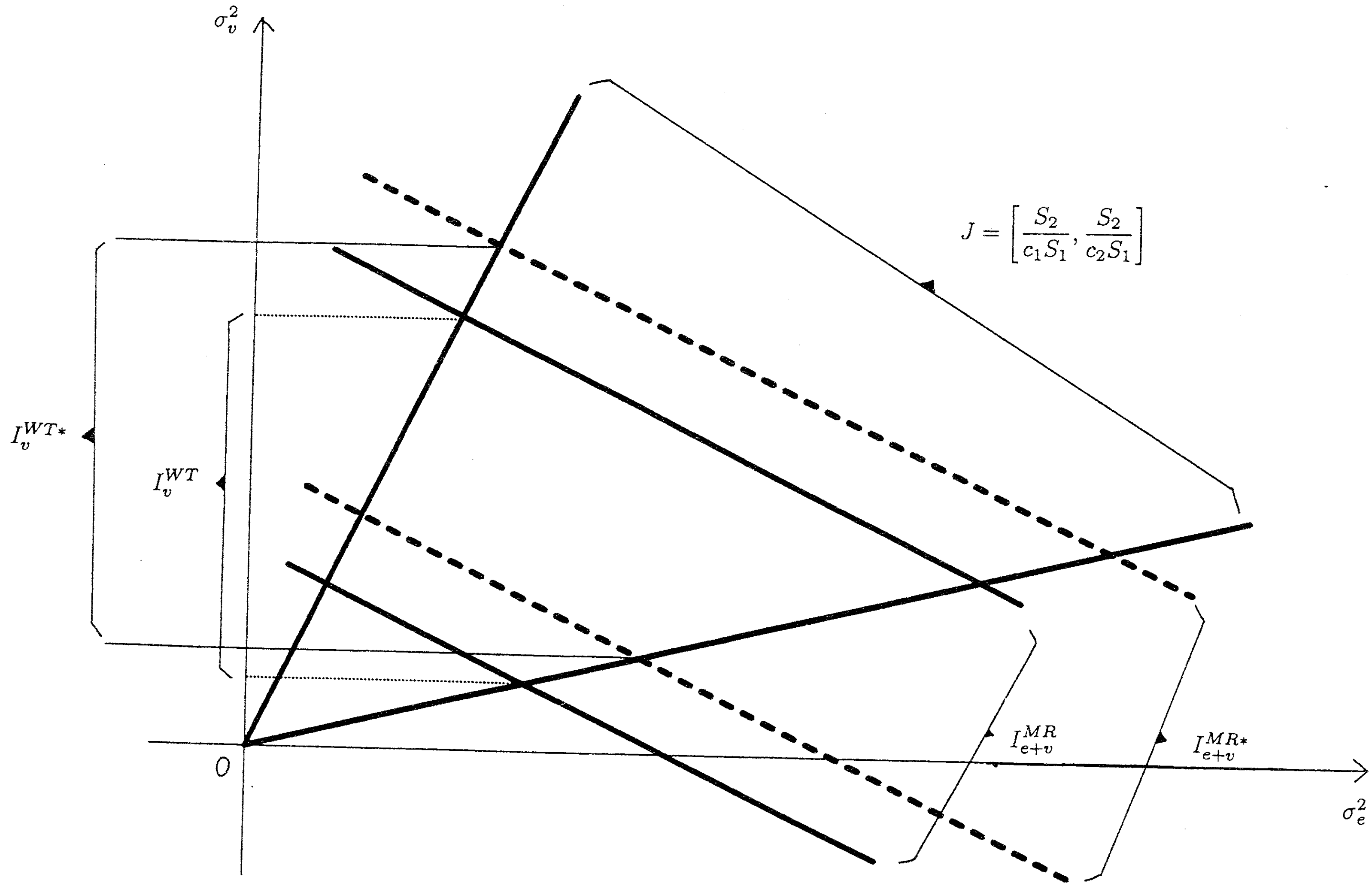

Figure 1. Derivation of the confidence intervals I_v^{WT} and I_v^{WT*}

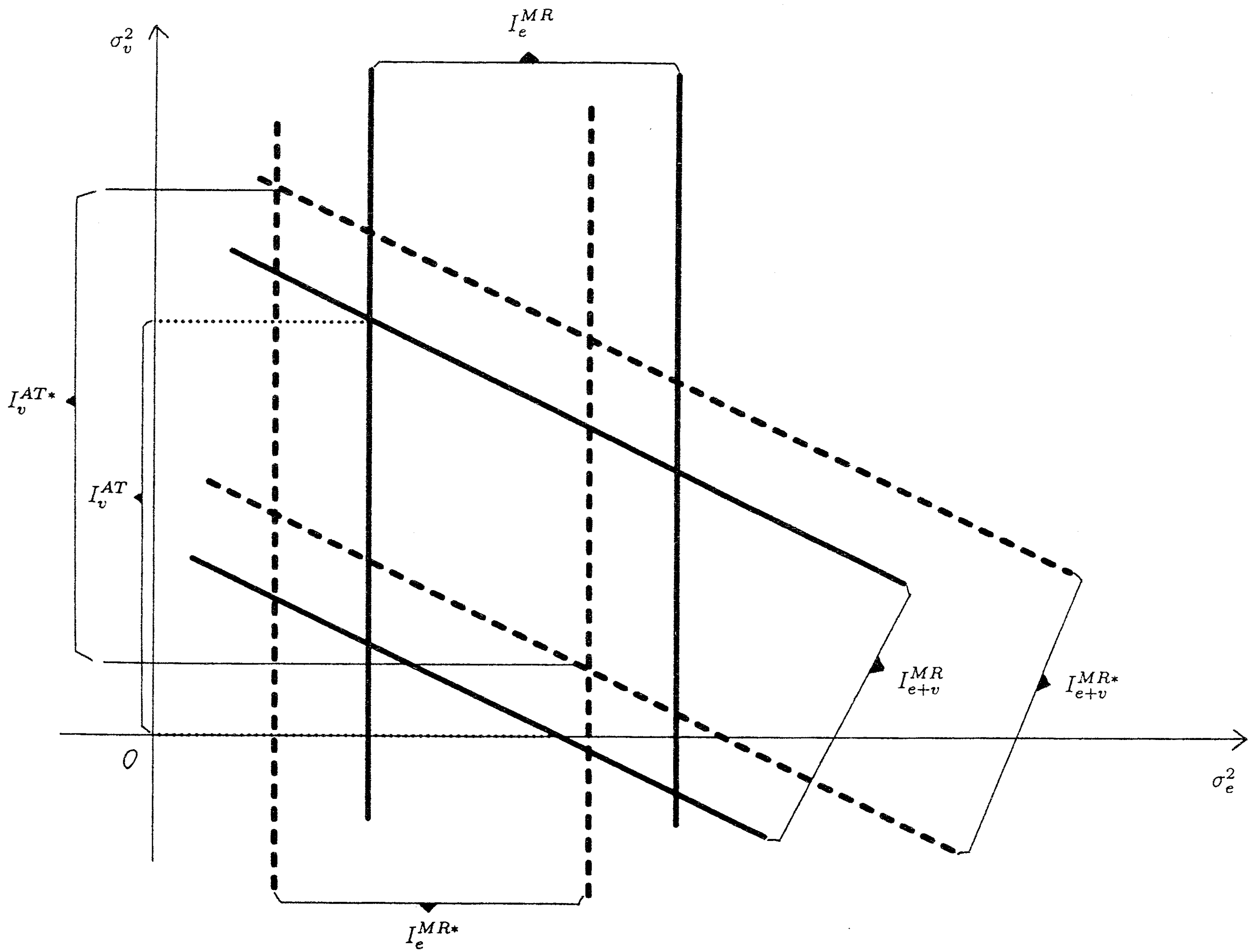

Figure 2. Derivation of the confidence intervals I_v^{AT} and I_v^{AT*}