The Annals of Statistics
1986, Vol. 14, No. 2, 679–690

# AN APPROACH TO UPPER BOUND PROBLEMS FOR RISKS OF GENERALIZED LEAST SQUARES ESTIMATORS

BY YASUYUKI TOYOOKA AND TAKEAKI KARIYA

*Osaka University and Hitotsubashi University*

First, an approach to an upper bound for the risk matrix of GLSE's is established when the information on the parameter space of the structural parameter in the covariance matrix of the error can be utilized. Second, this result is applied to regression with (i) serial correlation and (ii) heteroscedastic covariance structure. In the heteroscedastic regression, the problem of estimating the common mean of two normal populations is studied in detail.

**1. Introduction.** As is well known, in the regression model

$$y = X\beta + u, \qquad E(u) = 0 \quad \text{and} \quad \mathrm{Cov}(u) = \sigma^2\Sigma, \tag{1.1}$$

the Gauss–Markov estimator

$$\hat{\beta}(\Sigma) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \tag{1.2}$$

is the best linear unbiased estimator of $\beta$, provided $\Sigma$ is known. Here $X$ is an $n \times k$ fixed matrix of rank $k$ and $\Sigma$ is a positive definite matrix. Often, however, $\Sigma$ is a function of an estimable parameter, say $\Sigma = \Sigma(\theta)$, so that $\Sigma$ can and must be estimated based on $y$. In such a model, $\Sigma$ in (1.2) is replaced by an estimator, say $\hat{\Sigma} = \Sigma(\hat{\theta})$, and estimators of the form $\hat{\beta}(\hat{\Sigma})$ are often used in practice [see Theil (1971), Chapter 6]. In this article an estimator of this form shall be called a generalized least squares estimator (GLSE).

In many applications, the estimator for $\theta$ is based on the ordinary least squares (OLS) residual,

$$e = Ny = Nu \quad \text{with } N = I - X(X'X)^{-1}X'. \tag{1.3}$$

In Kariya and Toyooka (1985), when $\hat{\Sigma} = \Sigma(\hat{\theta}(e))$ and when the density function of $u$ belongs to a class of spherical density functions with mean 0 and covariance $\sigma^2\Sigma$,

$$f(u) = |\sigma^2\Sigma|^{-1/2} q\left(u'(\sigma^2\Sigma)^{-1}u\right)$$

(where the class is denoted as $S_n(0, \sigma^2\Sigma)$ below), it was shown that the risk matrix of $\hat{\beta}(\hat{\Sigma})$ is bounded below by the covariance matrix of $\hat{\beta}(\Sigma)$:

$$R(\hat{\beta}(\hat{\Sigma})) \equiv E[\hat{\beta}(\hat{\Sigma}) - \beta][\hat{\beta}(\hat{\Sigma}) - \beta]' \geq \mathrm{Cov}(\hat{\beta}(\Sigma)) = \sigma^2(X'\Sigma^{-1}X)^{-1}, \tag{1.4}$$

where throughout this article, inequalities for matrices should be understood in terms of nonnegative definiteness. Moreover, it is shown that (1.4) is valid for the

Received December 1983; revised March 1985.
*AMS* 1980 *subject classifications*. Primary 62J05; secondary 62M10.
*Key words and phrases*. GLSE, heteroscedasticity, intraclass correlation, serial correlation, upper bound for risk.

class which contains $\hat{\beta}(\hat{\Sigma})$,

$$P_1 = \{\tilde{\beta} | \tilde{\beta} = C(e)y,\ C(e)\text{: } k \times n \text{ measurable function of } e \text{ such that } C(e)X = I \text{ and } E\|\tilde{\beta}\|^2 < \infty\},$$

where $\|a\|^2 = a'a$ for $a \in R^k$. On the other hand, the uniform bounds for the approximation to the p.d.f. and the c.d.f. of GLSE were given in Toyooka and Kariya (1983).

In this paper we consider the problem of evaluating an upper bound for risk matrix of a GLSE in (1.4) under normality of $u$. Kariya (1981) also derived upper bounds for the risk of some GLSE's in Zellner's SUR model and a heteroscedastic model. Our approach here is different, more systematic, and thus applicable to many regression models with complicated covariance structure. As discussed in Remark 2.2 and Remark 2.3, our resulting upper bound uses the fourth moment of the estimator for the structural parameter to preserve the magnitude of the order. On the other hand, the upper bound using the second moment does not preserve the magnitude of the order which is discussed in Remark 2.2. In Section 2 a general approach to the upper-bound problem is established and in Section 3 it is applied to regressions with (i) serial correlation and (ii) heteroscedastic covariance structure. In the heteroscedastic regression, we treat as a special case the problem of estimating the common mean of two normal populations and compare the upper bound with the exact variance [see, e.g., Khatri and Shah (1975) and Cohen and Sackrowitz (1974) for the problem].

**2. Upper bound for the risk matrix of GLSE.** Let

$$y = X\beta + u, \qquad E(u) = 0, \quad \text{and} \quad \mathrm{Cov}(u) = \sigma^2\Sigma(\theta), \tag{2.1}$$

where $X$ is a fixed $n \times k$ matrix of rank $(X) = k$ and $\theta \in \Theta$ (nonempty open) $\subset R^1$. Assume that $\Sigma(\theta)$ is nonsingular in $\Theta$ such that

$$\Sigma^{-1}(\theta) = I_n + \lambda_n(\theta)C, \tag{2.2}$$

where $\lambda_n(\theta)$ is a continuous function of $\theta$ defined on $\Theta$ into $R^1$. Let $B$ be an orthogonal matrix such that

$$B'CB = \mathrm{diag}\{d_1, d_2, \dots, d_n\} \equiv D \quad \text{with } d_1 \leq \cdots \leq d_n. \tag{2.3}$$

Using $B$ and rewriting (2.1) as $B'y = B'X\beta + B'u$, without loss of generality, we can assume

$$\Sigma^{-1}(\theta) = I_n + \lambda_n(\theta)D. \tag{2.4}$$

And we further assume that $d_i \geq 0$ for all $i$ and $d_n > 0$. Typical examples in which (2.4) is satisfied are the covariance structure of serial correlation, intraclass correlation, and heteroscedasticity. The parameter $\theta$ is to be estimated based on the OLS residual $e$ in (1.3), which is often possible. For $\lambda \equiv \lambda_n(\theta) \equiv \lambda(\theta)$ in (2.2) we shall consider an estimator of the form $\hat{\lambda} = \lambda(\hat{\theta})$, where we assume $\hat{\theta} \in \Theta$ so that $\hat{\lambda} \in \Lambda = \{\lambda | \lambda = \lambda(\theta),\ \theta \in \Theta\}$.

To state a main result in this section, we introduce some notation. Let $Z$ be

$$ZZ' = N, \qquad Z'Z = I_{n-k}, \quad \text{and} \quad Z'X = 0 \tag{2.5}$$

and fix it throughout the article. Let

$$(2.6)\quad \begin{cases} A = (X'\Sigma^{-1}X)^{-1}, & \overline{X} = \Sigma^{-1/2}XA^{1/2}, \\ \overline{Z} = \Sigma^{1/2}Z(Z'\Sigma Z)^{-1/2} & \text{and} \quad \Gamma = [\overline{X}, \overline{Z}]. \end{cases}$$

Then $\Gamma$ is an $n \times n$ orthogonal matrix. Define

$$(2.7)\quad \tilde{\tilde{u}} = \Gamma'\tilde{u} = \begin{bmatrix} \overline{X}'\tilde{u} \\ \overline{Z}'\tilde{u} \end{bmatrix} = \begin{bmatrix} \tilde{\tilde{u}}_1 \\ \tilde{\tilde{u}}_2 \end{bmatrix},$$

where $\tilde{u} = \Sigma^{-1/2}u$. Now from (1.2) with $\hat{\Sigma}^{-1} = I + \hat{\lambda}D$, the GLSE $\hat{\beta}(\hat{\Sigma})$ is expressed as

$$(2.8)\quad \begin{aligned} \hat{\beta}(\hat{\Sigma}) - \beta &= (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}u \\ &= A^{1/2}(\overline{X}'\tilde{\Sigma}^{-1}\overline{X})^{-1}\overline{X}'\tilde{\Sigma}^{-1}\tilde{u} \\ &= A^{1/2}\tilde{\tilde{u}}_1 + A^{1/2}(\overline{X}'\tilde{\Sigma}^{-1}\overline{X})^{-1}\overline{X}'\tilde{\Sigma}^{-1}\overline{Z}\tilde{\tilde{u}}_2 \\ &\equiv (I) + (II) \quad \text{(say)}, \end{aligned}$$

where

$$(2.9)\quad \tilde{\Sigma} = \Sigma^{-1/2}\Sigma(\hat{\theta})\Sigma^{-1/2}, \qquad \hat{\theta} = \theta(e), \quad \text{and} \quad e = Z(Z'\Sigma Z)^{1/2}\tilde{\tilde{u}}_2.$$

Note that the second term $(II)$ is a function of $\tilde{\tilde{u}}_2$ only.

LEMMA 2.1 (Kariya and Toyooka, 1985). *Let* $\tilde{\tilde{u}} \in S_n(0, I)$. *If the second moments of* $\hat{\beta}(\hat{\Sigma})$ *exist, then*

$$(2.10)\quad \begin{aligned} R = E[\hat{\beta}(\hat{\Sigma}) - \beta][\hat{\beta}(\hat{\Sigma}) - \beta]' &= \sigma^2 A + A^{1/2}E[\Delta\Delta']A^{1/2} \\ &= \mathrm{Cov}(\hat{\beta}(\Sigma)) + E[\hat{\beta}(\hat{\Sigma}) - \hat{\beta}(\Sigma)][\hat{\beta}(\hat{\Sigma}) - \hat{\beta}(\Sigma)]', \end{aligned}$$

*where* $\Delta = (\overline{X}'\tilde{\Sigma}^{-1}\overline{X})^{-1}\overline{X}'\tilde{\Sigma}^{-1}\overline{Z}\tilde{\tilde{u}}_2$.

First it is remarked that this result is not restricted to the case $\Sigma(\theta)^{-1} = I + \lambda(\theta)D$ but it holds for any form of $\Sigma$. Second, a sufficient condition for the existence of the second moments of $\hat{\beta}(\hat{\Sigma}) - \beta$ is that $\hat{\theta} = \theta(e)$ is continuous and scale invariant, i.e., $\theta(e) = \theta(ae)$ for $a \in R^1$ [see Kariya and Toyooka (1985)]. Third, in the decomposition of the risk matrix in (2.10), the first term is the risk of the Gauss–Markov estimator and the second term is the loss of efficiency due to the estimation of $\theta$.

The evaluation of the loss $A^{1/2}E[\Delta\Delta']A^{1/2}$ is our concern. To do so, let

$$(2.11)\quad \begin{cases} B_1 = \{\hat{\lambda} - \lambda \geq 0\}, & B_2 = \{\hat{\lambda} - \lambda < 0\}, \\ W_1 \equiv 1, & W_2 = (1 + \lambda d_n)^2/(1 + \hat{\lambda}d_n)^2, \\ F = \mathrm{diag}\{d_1/(1 + \lambda d_1), \ldots, d_n/(1 + \lambda d_n)\} & \text{and} \quad L = \overline{X}'F\overline{Z}. \end{cases}$$

**Theorem** 2.1. *Assume that* $\Sigma^{-1}(\theta)$ *is of the covariance structure* (2.4) *and* $\hat{\theta} \in \Theta$ *(a.e.). Then*

$$A^{1/2}E[\Delta\Delta']A^{1/2} \le (g_1 + g_2)A, \tag{2.12}$$

*where* $g_i = E[\chi_{B_i}(\hat{\lambda} - \lambda)^2 W_i \tilde{\tilde{u}}_2' L'L\tilde{\tilde{u}}_2]$ $(i = 1, 2)$.

**Proof.** Since $\tilde{\Sigma} = \Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2}$,

$$\Delta = (\overline{X}'H\overline{X})^{-1}\overline{X}'H\overline{Z}\tilde{\tilde{u}}_2 \quad \text{with } H = \tilde{\Sigma}^{-1} = \Sigma^{1/2}\hat{\Sigma}^{-1}\Sigma^{1/2}. \tag{2.13}$$

Then

$$\begin{aligned} H &= \operatorname{diag}\left(\frac{1 + \hat{\lambda}d_1}{1 + \lambda d_1}, \dots, \frac{1 + \hat{\lambda}d_n}{1 + \lambda d_n}\right) \\ &= I_n + (\hat{\lambda} - \lambda)F. \end{aligned}$$

From (2.13) and $\overline{X}'\overline{Z} = 0$,

$$\Delta = \{I_k + (\hat{\lambda} - \lambda)\overline{X}'F\overline{X}\}^{-1}(\hat{\lambda} - \lambda)\overline{X}'F\overline{Z}\tilde{\tilde{u}}_2.$$

Let $J = I_k + (\hat{\lambda} - \lambda)\overline{X}'F\overline{X}$. Then

$$\Delta = (\hat{\lambda} - \lambda)J^{-1}L\tilde{\tilde{u}}_2. \tag{2.14}$$

Now, for any $a \in R^k$, by Schwarz's inequality,

$$\begin{aligned} a'E[\Delta\Delta']a &= E\left[(\hat{\lambda} - \lambda)a'J^{-1}L\tilde{\tilde{u}}_2\right]^2 \\ &\le E\left[(\hat{\lambda} - \lambda)^2 a'J^{-2}a\tilde{\tilde{u}}_2'L'L\tilde{\tilde{u}}_2\right]^2. \end{aligned} \tag{2.15}$$

Since $J$ depends on the sign of $(\hat{\lambda} - \lambda)$, we evaluate it on each $B_i$ in (2.11). On $B_1 = \{\hat{\lambda} - \lambda \geqq 0\}$

$$J = I_k + (\hat{\lambda} - \lambda)\overline{X}'F\overline{X} \geqq I_k$$

since $\overline{X}'F\overline{X} \geqq 0$. Then with $W_1 \equiv 1$ in (2.11)

$$a'J^{-2}a \leqq a'a = W_1 a'a. \tag{2.16}$$

Next, let

$$w = \max_{1 \leqq i \leqq n} \{d_i/(1 + \lambda d_i)\} = d_n/(1 + \lambda d_n), \tag{2.17}$$

since $f(x) = x/(1 + \lambda x)$ $(x \geqq 0)$ is increasing in $x$ for any $\lambda$. Then on $B_2$,

$$J \geqq I_k + (\hat{\lambda} - \lambda)wI_k = \{1 + (\hat{\lambda} - \lambda)w\}I_k,$$

whence with $W_2 = (1 + \lambda d_n)^2/(1 + \hat{\lambda}d_n)^2$ in (2.11),

$$a'J^{-2}a \leqq \{1 + (\hat{\lambda} - \lambda)w\}^{-2}a'a = W_2 a'a. \tag{2.18}$$

Therefore from (2.16) and (2.18),

$$a'E[\Delta\Delta']a \leqq \sum_{i=1}^{2} E\left[\chi_{B_i}(\hat{\lambda} - \lambda)^2 W_i \tilde{u}_2' L'L\tilde{u}_2\right] a'a \tag{2.19}$$

$$= (g_1 + g_2)a'a.$$

Thus, the desired result is obtained. □

REMARK 2.1. In the original term, $\tilde{u}_2'L'L\tilde{u}_2$ is expressed as

$$\tilde{u}_2'L'L\tilde{u}_2 = y'\left[I - \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\right]G\left[I - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\right]y,$$

where $G \equiv \Sigma^{-1/2}F\Sigma^{-1/2}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1/2}F\Sigma^{-1/2}$.

To evaluate the upper bound in (2.12) further, we assume that

$$\tilde{u} \sim N(0, \sigma^2 I_n).$$

LEMMA 2.2. *When $v \sim N(0, \sigma^2 I_m)$ and $C$ is an $m \times m$ matrix,*

$$E\left[(v'Cv)^2\right] = \sigma^4\left[(\operatorname{tr} C)^2 + \operatorname{tr} C'C + \operatorname{tr} C^2\right]. \tag{2.20}$$

Using this lemma with $C = L'L$ and applying Schwarz's inequality to $g_i$ in (2.12) yield

$$g_i \leqq \sigma^2\left[E\left\{\chi_{B_i}W_i^2(\hat{\lambda} - \lambda)^4\right\}\right]^{1/2}\left[(\operatorname{tr} L'L)^2 + 2\operatorname{tr}(LL')^2\right]^{1/2} \qquad (i = 1,2). \tag{2.21}$$

Now, combining (2.21) with (2.12) and (2.10), we obtain

THEOREM 2.2. *For $\Sigma$ in* (2.4),

$$\sigma^2 A \leqq R = R(\hat{\beta}(\hat{\Sigma})) \leqq \sigma^2 A + \delta(\bar{g}_1 + \bar{g}_2)\sigma^2 A, \tag{2.22}$$

*where* $\delta = [(\operatorname{tr} LL')^2 + 2\operatorname{tr}(LL')^2]^{1/2}$ *and* $\bar{g}_i = [E\{\chi_{B_i}W_i^2(\hat{\lambda} - \lambda)^4\}]^{1/2}$ $(i = 1, 2)$.

If $\hat{\theta}(e)$ is an even function of $e$, $\hat{\beta}(\hat{\Sigma})$ is unbiased for $\beta$ in which case $R(\hat{\beta}(\hat{\Sigma})) = \operatorname{Cov}(\hat{\beta}(\hat{\Sigma}))$.

A difficulty here is the evaluation of $\bar{g}_1$ and $\bar{g}_2$, which may be replaced by $[E\{(\hat{\lambda} - \lambda)^4\}]^{1/2}$ and $[E\{W_2^2(\hat{\lambda} - \lambda)^4\}]^{1/2}$, respectively, or may be both replaced by $[E\{(\hat{\lambda} - \lambda)^4\}]^{1/2}$.

REMARK 2.2. As is discussed in Section 3, $\hat{\theta} = \theta(e)$ with $e = Z(Z'\Sigma Z)^{1/2}\tilde{u}_2$ in (2.9) is often assumed to be $\hat{\theta} = \theta(ae) = \theta(e)$ for $a \in R^1$. So we assume $\hat{\theta} = \theta(\tilde{u}_2/\|\tilde{u}_2\|)$. Therefore, $\hat{\lambda} = \lambda(\tilde{u}_2/\|\tilde{u}_2\|)$. In this case, another evaluation for (2.15) is possible via Schwarz's inequality, i.e., for any $a \in R^k$,

$$a'E[\Delta\Delta']a \leq (n-k)\sum_{i=1}^{2} E\left[(\hat{\lambda} - \lambda)^2 W_i\right] a'a\delta_1,$$

where $\delta_1 = \operatorname{tr} L'L$. However, the r.h.s. of the above inequality is generally $O(1)$ as

$n \to \infty$ since $E[(\hat{\lambda} - \lambda)^2 W_t] = O(1/n)$. Therefore, this upper bound does not preserve the magnitude of the order as compared to the upper bound obtained in our theorem. See also Remark 2.3.

REMARK 2.3. The inequality (2.22) holds for any regressor $X$ and any estimator $\hat{\lambda} = \lambda(\hat{\theta})$. Further, the second term of the r.h.s. of (2.22) is of higher order in general. To see this, assume that $A = (X'\Sigma^{-1}X)^{-1} = O(1/n)$ or $\lim_{n\to\infty} X'\Sigma^{-1}X/n > 0$ exists as usual, and $\hat{\theta} - \theta = O_p(1/\sqrt{n})$. Then $\bar{g}_1 \leqq E[(\hat{\lambda} - \lambda)^4]^{1/2} = O(1/n)$ and from (2.18) $\bar{g}_2 \leqq E[\{W_2^2(\hat{\lambda} - \lambda)^4\}]^{1/2} = O(1/n)$ at least under such a condition as the boundedness of $W_2$. Hence, since $\delta = O(1)$ as is easily seen from the definitions of $F$ and $\delta$, $\delta(\bar{g}_1 + \bar{g}_2)(\sigma^2 A) = O(1/n^2)$, while $\sigma^2 A = O(1/n)$. Therefore, our upper bound (2.22) preserves the order structure, contrary to the case in Remark 2.2.

REMARK 2.4. It is interesting to see that if $\bar{X}$ is of the form $P\begin{bmatrix} 0 \\ I_k \end{bmatrix}Q$ with permutation matrices $P$ and $Q$, $\delta = 0$, because

$$LL' = \bar{X}'F\bar{Z}\bar{Z}'F\bar{X} = \bar{X}'F[I - \bar{X}\bar{X}']F\bar{X} = \bar{X}'F^2\bar{X} - (\bar{X}'F\bar{X})^2$$

implies $\operatorname{tr} LL' = 0$ and $\operatorname{tr}(LL')^2 \leqq (\operatorname{tr} LL')^2 = 0$. From Theorem 2.2 this implies that the GLSE is as efficient as the Gauss–Markov estimator. In fact, the following proposition holds in general.

PROPOSITION 2.1. *$\delta = 0$ if and only if $X'\Sigma Z = 0$.*

PROOF. $\delta = 0$ implies $\operatorname{tr} LL' = 0$, which in turn implies $L = \bar{X}'F\bar{Z} = 0$. Hence, from the definition of $F$, $\bar{X}$, and $\bar{Z}$, $\bar{X}'(I - \lambda F)\bar{Z} = 0$ or $\bar{X}\Sigma Z = 0(X'\Sigma Z = 0)$. Tracing back this proof yields the converse, completing the proof. □

The condition $X'\Sigma Z = 0$ in Proposition 2.1 is well known as a necessary and sufficient condition for the OLSE $\hat{\beta}(I)$ to be identically equal to the Gauss–Markov estimator $\hat{\beta}(\Sigma)$ [see, e.g., Rao (1967)]. Hence, under this condition, $\hat{\beta}(I) = \hat{\beta}(\Sigma) = \hat{\beta}(\hat{\Sigma})$ and so $\operatorname{Cov}(\hat{\beta}(\Sigma)) = \operatorname{Cov}(\hat{\beta}(\hat{\Sigma}))$. Theorem 2.2 together with Proposition 2.1 provides an alternative proof of the above result by Rao.

**3. Applications.** In this section the results obtained in Section 2 are applied to two cases, regressions with (i) serial correlation and (ii) heteroscedastic covariance structure.

EXAMPLE 1 (Regression with serial correlation). We consider the model (2.1) with errors of first-order serial correlation:

$$u_t = \theta u_{t-1} + \varepsilon_t, \qquad \theta \in \Theta = \{\theta : |\theta| < 1\}, \tag{3.1}$$

where $u = (u_1, \ldots, u_n)'$ and $\{\varepsilon_t\} \sim$ i.i.d. $N(0, \sigma_\varepsilon^2)$. As is well known, the inverse matrix of the covariance matrix $V$ of $u$ is given by

$$(3.2)\qquad V^{-1} = \frac{1}{\sigma_\epsilon^2}\begin{bmatrix} 1 & -\theta & & & 0 \\ -\theta & 1+\theta^2 & & \ddots & \\ & & \ddots & & \\ & \ddots & & 1+\theta^2 & -\theta \\ 0 & & & -\theta & 1 \end{bmatrix}.$$

This matrix is often approximated by the matrix $W^{-1}$ in which the (1, 1)th and $(n, n)$th elements of $V^{-1}$ are replaced by $1-\theta+\theta^2$. Then $W^{-1} = [(1-\theta)^2 I + \theta C]/\sigma_\epsilon^2$, where

$$(3.3)\qquad C = \begin{bmatrix} 1 & -1 & & & 0 \\ -1 & 2 & & \ddots & \\ & & \ddots & & \\ & \ddots & & 2 & -1 \\ 0 & & & -1 & 1 \end{bmatrix}.$$

Using the scale invariance of GLSE $\hat{\beta}(\hat{\Sigma})$, regard $\Sigma$ in (2.1) as the inverse of

$$(3.4)\qquad \Sigma^{-1} = I_n + \lambda(\theta)C \quad \text{with } \lambda = \theta/(1-\theta)^2,$$

where $\sigma^2 = \sigma_\epsilon^2/(1-\theta^2)$. Note $-\frac{1}{4} < \lambda < \infty$ from $|\theta| < 1$. As in Durbin and Watson (1971), the eigenvalues of $C$ are given by

$$(3.5)\qquad d_j = 2[1 - \cos(\pi(j-1)/n)] \qquad (j = 1, \dots, n).$$

It is noted that $0 \leqq d_j < \max_{1 \leqq j \leqq n} d_j = d_n < 4$ and $d_j \neq 0$ except $d_1$ and that $d_1 = 0 < d_2 < \cdots < d_n$. So we can assume by (2.4)

$$(3.6)\qquad \Sigma^{-1} = I_n + \lambda(\theta)D \quad \text{with } D = \operatorname{diag}(0, d_2, \dots, d_n).$$

Next, choose an estimator $\hat{\theta}$ of $\theta$ based on the OLS residual $e$ such that

$$(3.7)\qquad \begin{cases} |\hat{\theta}(e)| < 1, \quad \hat{\theta}(-e) = \hat{\theta}(e), \quad \hat{\theta}(ae) = \hat{\theta}(e) \quad \text{for } a > 0 \text{ and} \\ \hat{\theta} \text{ is continuous.} \end{cases}$$

A typical choice will be

$$(3.8)\qquad \hat{\theta}_0 = \sum_{t=2}^{n} e_t e_{t-1} \Big/ \sum_{t=1}^{n} e_t^2 = e'Ke/e'e,$$

where $e = (e_1, e_2, \dots, e_n)'$ and $K = (k_{ij})$ with $k_{ij} = 0$ except $k_{i,i+1} = k_{j+1,1} = \frac{1}{2}$ $(j = 2, \dots, n,\ i = 1, \dots, n-1)$. Note that an application of Schwarz's inequality shows $|\hat{\theta}_0| < 1$. The second and third conditions of (3.7) guarantee the unbiasedness and the existence of the second moment of the GLSE $\hat{\beta}(\hat{\Sigma})$, respectively. Define

$$(3.9)\qquad \hat{\lambda} = \hat{\theta}/(1-\hat{\theta})^2 \quad \text{and} \quad \hat{\Sigma} = (I + \hat{\lambda}D)^{-1}.$$

Now we derive an upper bound for $\operatorname{Cov}(\hat{\beta}(\hat{\Sigma}))$. Let $\Theta = \{\theta \colon |\theta| < 1\}$. Since $\hat{\theta} \in \Theta$, the inequality (2.22) in Theorem 2.2 is valid.

For the practical use it is necessary to evaluate $\sigma^2 A + \delta(\bar{g}_1 + \bar{g}_2)\sigma^2 A$ in the inequality exactly. But here we obtain an approximation to this.

LEMMA 3.1. *Assume* $\hat{\theta} - \theta = O_p(1/\sqrt{n})$. *Then*

$$W_2 = 1 + O_p(1/\sqrt{n}).$$

PROOF. For the evaluation of $W_2$, we consider the behaviour of

$$(1 + \lambda d_n)/(1 + \hat{\lambda} d_n). \tag{3.10}$$

Expanding $\hat{\lambda} = \lambda(\hat{\theta})$ around $\theta$,

$$\begin{aligned}\hat{\lambda} &= \lambda(\theta) + (\hat{\theta} - \theta)\frac{d}{d\theta}\lambda(\theta)\Big|_{\theta=\theta^*} \quad \text{for } \theta < \theta^* < \hat{\theta} \\ &= \lambda(\theta) + O_p(1/\sqrt{n}).\end{aligned} \tag{3.11}$$

Remark that by Taylor's expansion,

$$\begin{aligned} d_n &= 2\left[1 - \cos\left(\pi\frac{n-1}{n}\right)\right] \\ &= 4 + \frac{2\pi}{n} + o(1/n).\end{aligned} \tag{3.12}$$

Put (3.11) and (3.12) into (3.10),

$$\begin{aligned}\frac{1 + \lambda d_n}{1 + \hat{\lambda} d_n} &= \frac{\alpha + 2\pi\lambda/n}{\alpha + 2\pi\lambda/n + O_p(1/\sqrt{n})} \\ &= 1 + O_p(1/\sqrt{n}) \quad \text{with } \alpha = 1 + 4\lambda > 0,\end{aligned}$$

which completes the proof. □

On the other hand,

$$\bar{g}_i \leqq \left[E\left\{W_i^2(\hat{\lambda} - \lambda)^4\right\}\right]^{1/2} \qquad (i = 1,2). \tag{3.13}$$

Using Lemma 3.1, for the typical choice $\hat{\theta}_0$ in (3.8), the leading terms of the r.h.s. of (3.13) are evaluated by the usual $\delta$-method as

LEMMA 3.2. *For the typical choice* $\hat{\theta}_0$, *the leading terms of the r.h.s. of* (3.13) *are both evaluated as*

$$\left[E\left\{W_i^2(\hat{\lambda} - \lambda)^4\right\}\right]^{1/2} = \frac{1}{n}\frac{\sqrt{3}(1+\theta)}{(1-\theta)^7} + o(1/n) \qquad (i = 1,2).$$

PROOF. Since $d/d\theta\, \lambda(\theta) = (1 + \theta)/(1 - \theta)^3$, $\lambda(\hat{\theta}_0) - \lambda(\theta)$ is approximately $(\hat{\theta}_0 - \theta)(1 + \theta)/(1 - \theta)^3$. By Lemma 3.1 $W_i^2(\hat{\lambda} - \lambda)^4$ is asymptotically $(\hat{\theta}_0 - \theta)^4(1 + \theta)^4/(1 - \theta)^{12}$. Remark that $\sqrt{n}(\hat{\theta}_0 - \theta)$ converges to $N(0, 1/(1 - \theta^2))$ in distribution. Then by using the fourth-order central moment of a normal random variable, the result is obtained. □

Since

$$\sigma^2 A + \delta(\bar{g}_1 + \bar{g}_2)\sigma^2 A \leqq \sigma^2 A + \delta(\bar{h}_1 + \bar{h}_2)\sigma^2 A, \tag{3.14}$$

where $\bar{h}_i = [E\{W_i^2(\hat{\lambda} - \lambda)^4\}]^{1/2}$ $(i = 1, 2)$, the leading term of the r.h.s. of (3.14) is obtained as

THEOREM 3.1. *For the choice* $\hat{\theta}_0$, *an approximation to the r.h.s. of* (3.14) *is*

$$\sigma^2 A + 2\delta \frac{1}{n} \frac{\sqrt{3}(1+\theta)}{(1-\theta)^7} \sigma^2 A.$$

EXAMPLE 2 (Regression with heteroscedastic covariances). The heteroscedastic covariance structure in (2.1) is given by

$$\mathrm{Cov}(u) = \begin{bmatrix} \theta_1 I_{n_1} & 0 \\ 0 & \theta_2 I_{n_2} \end{bmatrix} \quad (\theta_1 > 0,\ \theta_2 > 0). \tag{3.15}$$

For the scale invariance of GLSE, without loss of generality, assume

$$\Sigma^{-1}(\theta) = \begin{bmatrix} I_{n_1} & 0 \\ 0 & \theta I_{n_2} \end{bmatrix} \quad \text{with } \theta = \theta_1/\theta_2. \tag{3.16}$$

So the parameter space is $\Theta = \{\theta \colon \theta > 0\}$. Then

$$\Sigma^{-1}(\theta) = I_n + \lambda(\theta) D, \tag{3.17}$$

where $\lambda(\theta) = \theta - 1$ and

$$D = \mathrm{diag}\Big(\overbrace{0, \dots, 0}^{n_1}, \overbrace{1, \dots, 1}^{n_2}\Big).$$

Hence from the inequalities (2.12) in Theorem 2.1 and (2.22) in Theorem 2.2, if $\hat{\theta} \in \Theta$ with probability one,

$$\begin{aligned} \sigma^2 A \le R(\hat{\beta}(\hat{\Sigma})) &\le \left\{1 + \sum_{i=1}^{2} E\left[\chi_{B_i}(\hat{\theta} - \theta)^2 W_i \tilde{\tilde{u}}_2' L' L \tilde{\tilde{u}}_2\right]\right\} \sigma^2 A \\ &\le \left\{1 + \sum_{i=1}^{2} \delta\left[E\left\{\chi_{B_i} W_i^2 (\hat{\theta} - \theta)^4\right\}\right]^{1/2}\right\} \sigma^2 A, \end{aligned} \tag{3.18}$$

where $\sigma^2 = \theta_1$, $B_1 = \{\hat{\theta} - \theta \ge 0\}$, $B_2 = \{\hat{\theta} - \theta < 0\}$, $W_1 = 1$, and $W_2 = (\theta/\hat{\theta})^2$. Here the expression of $\tilde{\tilde{u}}_2' L' L \tilde{\tilde{u}}_2$ in the original term is given in Remark 2.1. When $\theta = \theta_1/\theta_2 \ge 1$ is assumed so that $\Theta_1 = \{\theta \ge 1\}$, a further evaluation is given as follows.

THEOREM 3.2. *For $\theta \geq 1$, if $\hat{\theta} \in \Theta_1$ with probability one,*

$$\sigma^2 A \leq R(\hat{\beta}(\hat{\Sigma})) \leq \left\{1 + (1+\theta^2) E\left[(\hat{\theta}-\theta)^2 \tilde{\tilde{u}}_2' L' L \tilde{\tilde{u}}_2\right]\right\}\sigma^2 A$$
$$\leq \left\{1 + (1+\theta^2)\,\delta\left[E(\hat{\theta}-\theta)^4\right]^{1/2}\right\}\sigma^2 A.$$

A natural estimator for $\theta$ is proposed as follows: let

$$\hat{\theta}_i = \frac{1}{n_i - k} e' C_i e \qquad (i = 1, 2)$$

with

$$C_1 = \mathrm{diag}\Big(\overbrace{1, 1, \dots, 1}^{n_1}, \overbrace{0, \dots, 0}^{n_2}\Big)$$

and

$$C_2 = \mathrm{diag}\Big(\overbrace{0, \dots, 0}^{n_1}, \overbrace{1, \dots, 1}^{n_2}\Big)$$

and let

$$\hat{\theta}_0 = \max(1, \hat{\theta}_1/\hat{\theta}_2).$$

In this case, $\hat{\theta}_0$ is a continuous, scale invariant, and even function of $e$. Then $\hat{\beta}(\hat{\Sigma})$ is unbiased and the second moment of $\hat{\beta}(\hat{\Sigma})$ exists. Therefore, $R(\hat{\beta}(\hat{\Sigma})) = \mathrm{Cov}(\hat{\beta}(\hat{\Sigma}))$.

As a special case, let us consider the problem of estimating the common mean of two normal populations and compare the upper bound for $R(\hat{\beta}(\hat{\Sigma}))$ in (3.18) with the exact variance (risk). The problem has been treated by many authors. Among others, Graybill and Deal (1959) raised the problem and proposed a GLS type estimator, Cohen and Sackrowitz (1974) proposed different estimators using an ancillary statistic, and Khatri and Shah (1975) presented methods for evaluating the exact variance of an estimator encountered in such a model. In our context, the model is given by $y = X\beta + u$, where $\mathrm{Cov}(u) = \theta_1 \Sigma(\theta)$ is given by (3.15) and (3.16) and

$$X = \mathbf{1} = (1, \dots, 1)' \colon (n_1 + n_2) \times 1.$$

Then a GLSE $\hat{\beta}(\hat{\Sigma})$ here is evaluated as

$$\hat{\beta}(\hat{\Sigma}) = \left[\frac{n_1 \bar{y}_1}{\hat{\theta}_1} + \frac{n_2 \bar{y}_2}{\hat{\theta}_2}\right] \Big/ \left[\frac{n_1}{\hat{\theta}_1} + \frac{n_2}{\hat{\theta}_2}\right], \tag{3.19}$$

where $\bar{y}_1 = \sum_{i=1}^{n_1} y_i/n_1$ and $\bar{y}_2 = \sum_{i=n_1+1}^{n_1+n_2} y_i/n_2$ with $y = (y_i)$, and $\hat{\theta}_i \equiv \hat{\theta}_i(e)$ is assumed to satisfy $\hat{\theta}_i(-e) = \hat{\theta}_i(e)$ and $\hat{\theta}_i(ae) = \hat{\theta}_i(e)$ for $a > 0$ $(i = 1, 2)$. First, we shall evaluate the upper bound in (3.18). From Remark 2.1 and (3.16), it is easily shown that $\tilde{\tilde{u}}_2' L' L \tilde{\tilde{u}}_2$ is evaluated as

$$\tilde{\tilde{u}}_2' L' L \tilde{\tilde{u}}_2 = n_1^2 n_2^2 (\bar{y}_2 - \bar{y}_1)^2 / \theta_1 (n_1 + n_2\theta)^3.$$

Hence from (3.18) the upper bound for the covariance matrix $R(\hat{\beta}(\hat{\Sigma}))$ is given by

$$\left\{1 + E\left[\frac{n_1^2 n_2^2(\hat{\theta} - \theta)^2(\bar{y}_2 - \bar{y}_1)^2}{\theta_1(n_1 + n_2\theta)^3}\left(\chi_{B_1} + \chi_{B_2}\left(\frac{\theta}{\hat{\theta}}\right)^2\right)\right]\right\}\theta_1 A, \tag{3.20}$$

where $A = (X'\Sigma^{-1}X)^{-1} = (n_1 + n_2\theta)^{-1}$. On the other hand, the exact variance of $\hat{\beta}(\hat{\Sigma})$ in (3.19) can be evaluated in a line with Khatri and Shah (1975). In particular, if $\hat{\theta}_i$ is a linear combination of $s_1^2$, $s_2^2$, and $(\bar{y}_1 - \bar{y}_2)^2$, we can use their result where $s_i^2$ is the sample unbiased variance of each population. However, even in this case it is difficult to analytically compare (3.20) with the exact variance provided by them. For this reason, we make use of Lemma 2.1 and compare them indirectly. Since from (3.18)

$$\begin{aligned}\hat{\beta}(\hat{\Sigma}) - \hat{\beta}(\Sigma) &= \frac{n_1\bar{y}_1 + n_2\bar{y}_2\hat{\theta}}{n_1 + n_2\hat{\theta}} - \frac{n_1\bar{y}_1 + n_2\bar{y}_2\theta}{n_1 + n_2\theta}\\ &= \frac{n_1 n_2(\hat{\theta} - \theta)(\bar{y}_2 - \bar{y}_1)}{(n_1 + n_2\hat{\theta})(n_1 + n_2\theta)},\end{aligned}$$

the exact variance of $\hat{\beta}(\hat{\Sigma})$ is obtained from Lemma 2.1 as

$$\operatorname{Cov}(\hat{\beta}(\hat{\Sigma})) = \left\{1 + E\left[\frac{n_1^2 n_2^2(\hat{\theta} - \theta)^2(\bar{y}_2 - \bar{y}_1)^2}{\theta_1(n_1 + n_2\theta)(n_1 + n_2\hat{\theta})^2}\right]\right\}\theta_1 A. \tag{3.21}$$

Therefore the difference between the upper bound (3.20) and the exact variance (3.21) is

$$\begin{aligned}\zeta &\equiv \theta_1 A E\left\{\frac{n_1^2 n_2^2(\hat{\theta} - \theta)^2(\bar{y}_2 - \bar{y}_1)^2}{\theta_1(n_1 + n_2\theta)^3}\right.\\ &\qquad\left.\times\left[\chi_{B_1} + \chi_{B_2}\left(\frac{\theta}{\hat{\theta}}\right)^2 - \left(\frac{n_1 + n_2\theta}{n_1 + n_2\hat{\theta}}\right)^2\right]\right\}\\ &= E\left\{n_1^2 n_2^2(\hat{\theta} - \theta)^2(\bar{y}_2 - \bar{y}_1)^2[(\hat{\theta} - \theta)Q]\right\}/(n_1 + n_2\theta)^4,\end{aligned} \tag{3.22}$$

where

$$Q = \frac{2n_2\hat{\theta}\chi_{B_1} + 2n_1}{\hat{\theta}(n_1 + n_2\hat{\theta})} + (\hat{\theta} - \theta)\left[\hat{\theta}^{-2}\chi_{B_2} - \left(\frac{n_2}{n_1 + n_2\hat{\theta}}\right)^2\right].$$

If $\hat{\theta} - \theta = O_p(m^{-1/2})$ with $m = \min(n_1, n_2)$, then the difference $\zeta$ of the upper bound and the exact variance is of order $O(m^{-5/2})$ under regularity conditions. Hence the $m^2\zeta = O(m^{-1/2})$ and thus the upper bound is asymptotically equivalent to the exact variance up to $O(m^{-2})$. When $\hat{\theta}_i$ is independent of $(\bar{y}_2 - \bar{y}_1)^2$ as is the case of $\hat{\theta}_i = s_i^2$ $(i = 1, 2)$, $(\bar{y}_2 - \bar{y}_1)^2$ can be replaced by $E(\bar{y}_2 - \bar{y}_1)^2 = (n_1 + n_2\theta)\theta_2/n_1 n_2$ in (3.20), (3.21), and (3.22). If $\hat{\theta}_i = s_i^2$, using $(n_1 - 1)s_1^2/(n_2 - 1)s_2^2 = \theta v/(1 - v)$ with $v$ being distributed as beta$((n_1 - 1)/2, (n_2 - 1)/2)$, we can also evaluate (3.21) exactly if necessary.

**Acknowledgments.** The authors are deeply grateful to the referees for their invaluable suggestions and comments. Portions of Kariya's work were supported by the Ministry of Education of Japan under General C59540104.

## REFERENCES

BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā* **15** 377–380.

COHEN, A. and SACKROWITZ, H. B. (1974). On estimating the common mean of two normal populations. *Ann. Statist.* **2** 1274–1282.

DURBIN, J. and WATSON, G. S. (1971). Testing for serial correlation in least squares regression III. *Biometrika* **58** 1–19.

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 2SAE in a heteroscedastic model. *J. Amer. Statist. Assoc.* **76** 975–979.

KARIYA, T. and TOYOOKA, Y. (1985). Nonlinear versions of the Gauss–Markov theorem and GLSE. In *Multivariate Analysis* (P. R. Krishnaiah, ed.) **6** 345–354. Elsevier, New York.

KHATRI, C. G. and SHAH, K. R. (1975). Exact variance of combined inter- and intra-block estimates in incomplete block designs. *J. Amer. Statist. Assoc.* **70** 402–406.

MCELROY, F. W. (1967). A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased. *J. Amer. Statist. Assoc.* **62** 1302–1304.

RAO, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 355–372. Univ. California Press.

THEIL, H. (1971). *Principles of Econometrics.* Wiley, New York.

TOYOOKA, Y. and KARIYA, T. (1983). Uniform bounds for approximations to the pdf's and cdf's of GLSP and GLSE. Discussion paper 84, Hitotsubashi Univ.

DEPARTMENT OF APPLIED MATHEMATICS
OSAKA UNIVERSITY
TOYONAKI-SHI, OSAKA 560
JAPAN

INSTITUTE OF ECONOMIC RESEARCH
HITOTSUBASHI UNIVERSITY
KUNITACHI-SHI, TOKYO 186
JAPAN