The Annals of Statistics
1996, Vol. 24, No. 4, 1547–1559

LEAST UPPER BOUND FOR THE COVARIANCE MATRIX OF A GENERALIZED LEAST SQUARES ESTIMATOR IN REGRESSION WITH APPLICATIONS TO A SEEMINGLY UNRELATED REGRESSION MODEL AND A HETEROSCEDASTIC MODEL

BY HIROSHI KURATA AND TAKEAKI KARIYA

Yamaguchi University and Hitotsubashi University

In a general normal regression model, this paper first derives the least upper bound (LUB) for the covariance matrix of a generalized least squares estimator (GLSE) relative to the covariance matrix of the Gauss–Markov estimator. Second the result is applied to the (unrestricted) Zellner estimator in an N-equation seemingly unrelated regression (SUR) model and to the GLSE in a heteroscedastic model.

1. Introduction. A normal regression model of the form

$$y = X\beta + \varepsilon \quad \text{with } \varepsilon \sim N(0, \Omega), \tag{1.1}$$

where

$$\Omega = \Omega(\theta) \in S(n) \tag{1.2}$$

is a typical model often used in applications. Here X is a regression matrix of $n \times k$, rank $X = k$, the covariance matrix in (1.2) of error term ε is a function of a $d \times 1$ unknown vector θ with $d < n$ and $S(n)$ is the set of $n \times n$ positive definite matrices. Models of the form (1.1) include, for example, the SUR model formulated by Zellner (1962, 1963) and the heteroscedastic model where Ω is diagonal. In the estimation of β in (1.1), a GLSE of the form

$$\hat{\beta}(\hat{\Omega}) = (X'\hat{\Omega}^{-1}X)^{-1} X'\hat{\Omega}^{-1}y \tag{1.3}$$

is often used where $\hat{\Omega} = \Omega(\hat{\theta})$ with $\hat{\theta}$ an estimator of θ. The problem we consider in this paper is to study the efficiency of a GLSE in (1.3) in terms of the covariance matrix $\mathrm{Cov}(\hat{\beta}(\hat{\Omega}))$ relative to the covariance matrix of what we call the Gauss–Markov estimator (GME) $\hat{\beta}(\Omega)$ even if Ω is unknown. As is well known, the GME

$$\hat{\beta}_{\mathrm{GME}} = \hat{\beta}(\Omega)$$

is the best linear unbiased estimator when Ω is known. However, the GLSE $\hat{\beta}(\hat{\Omega})$ in (1.3) is, in general, nonlinear in y and hence it is difficult to derive $\mathrm{Cov}(\hat{\beta}(\hat{\Omega}))$ explicitly and to study the efficiency of the GLSE $\hat{\beta}(\hat{\Omega})$ unless an

Received April 1994; revised March 1995.
AMS 1991 *subject classifications*. Primary 62J05; secondary 62M10.
Key words and phrases. Nonlinear Gauss-Markov theorem, efficiency of GLSE, seemingly unrelated equation, heteroscedastic model, Kantorovich inequality.

additional structure is available. In this paper, the condition we impose on a GLSE of the form (1.3) is the following distributional property: conditional on $\hat{\Omega}$,

$$\hat{\beta}(\hat{\Omega}) \text{ given } \hat{\Omega} \sim N(\beta, H), \tag{1.4}$$

where H is the conditional covariance matrix of $\hat{\beta}(\hat{\Omega})$ given by

$$H \equiv H(\hat{\Omega}, \Omega) = (X'\hat{\Omega}^{-1}X)^{-1} X'\hat{\Omega}^{-1}\Omega\hat{\Omega}^{-1}X(X'\hat{\Omega}^{-1}X)^{-1}. \tag{1.5}$$

For such a GLSE, the covariance matrix is expressed as

$$\mathrm{Cov}\big(\hat{\beta}(\hat{\Omega})\big) = E[H]. \tag{1.6}$$

A simple example of a GLSE satisfying (1.4) is the ordinary least squares estimator (OLSE) $\hat{\beta}(I)$ in which case

$$H_{\mathrm{OLSE}} = H(I, \Omega) = (X'X)^{-1} X'\Omega X(X'X)^{-1}$$

is nonrandom. For a GLSE satisfying the property (1.4), the covariance matrix (1.6) is bounded below by the covariance matrix of the GME $\hat{\beta}_{\mathrm{GME}} = \hat{\beta}(\Omega)$:

$$\mathrm{Cov}\big(\hat{\beta}(\hat{\Omega})\big) \geq \mathrm{Cov}\big(\hat{\beta}_{\mathrm{GME}}\big) = (X'\Omega^{-1}X)^{-1} \equiv A^{-1} \tag{1.7}$$

[see Kariya (1981) and Kariya and Toyooka (1985)], where inequalities for matrices here should be understood in terms of nonnegative definiteness. Hence an efficiency of a GLSE $\hat{\beta}(\hat{\Omega})$ that satisfies (1.4) is measured relative to the GME by such a quantity as

$$\eta = |\,\mathrm{Cov}\big(\hat{\beta}(\hat{\Omega})\big)\,| / |\,\mathrm{Cov}\big(\hat{\beta}_{\mathrm{GME}}\big)\,| \geq 1 \tag{1.8}$$

provided $E[H]$ is evaluated. However, it is still difficult to derive $E[H]$ explicitly since H in (1.5) is nonlinear in y, and even if it is derived, it is often very complicated. Consequently, to obtain useful information on the efficiency of a given GLSE satisfying (1.4), we may formulate our problem as the problem of finding an effective upper bound for the covariance matrix $\mathrm{Cov}(\hat{\beta}(\hat{\Omega})) = E[H]$ relative to its lower bound $\mathrm{Cov}(\hat{\beta}_{\mathrm{GME}}) = A^{-1}$:

$$E[H] \leq \alpha(\hat{\Omega}) A^{-1} \tag{1.9}$$

[see Kariya (1981) and Toyooka and Kariya (1986)]. Here $\alpha(\hat{\Omega})$ in (1.9) is a nonrandom scalar function associated with $H \equiv H(\hat{\Omega}, \Omega)$ in (1.5). Clearly such an upper bound $\alpha(\hat{\Omega})$ can be viewed as an upper bound for the efficiency of a GLSE in a strong sense. For example, for η in (1.8), (1.9) implies an upper bound for η:

$$1 \leq \eta \leq \big\{\alpha(\hat{\Omega})\big\}^k. \tag{1.10}$$

In Section 2, in order to obtain an effective upper bound in (1.9), we first derive the least upper bound (LUB) $a_l(\hat{\Omega})$ for the conditional covariance matrix H relative to the lower bound A^{-1},

$$H \leq a_l(\hat{\Omega}) A^{-1} \quad \text{a.e.}, \tag{1.11}$$

and then propose the expected LUB,

$$\alpha_l(\hat{\Omega}) = E\left[a_l(\hat{\Omega})\right], \tag{1.12}$$

as $\alpha(\hat{\Omega})$ in (1.9). However, in general it is still not easy to evaluate the expected value of $a_l(\hat{\Omega})$ in (1.12) except for certain cases. Hence we also derive a simpler but effective bound for H and hence for $E[H]$. These results are applicable to any given GLSE so long as it satisfies (1.4).

Typical GLSEs which satisfy (1.4) are the Zellner estimator in the SUR model and the GLSE in the heteroscedastic model. In Sections 3 and 4, applying the general results in Section 2 to these models, we derive the expected LUB and an effective upper bound for the covariance matrices in these models. More specifically, let the N-equation SUR model be expressed as a model in (1.1) with

$$\begin{aligned} &X = \operatorname{diag}\{X_1, \dots, X_N\}, \\ &y = (y_1', \dots, y_N')', \qquad \varepsilon = (\varepsilon_1', \dots, \varepsilon_N')', \\ &\beta = (\beta_1', \dots, \beta_N')', \qquad \Omega = \Sigma \otimes I_m \quad \text{with } \Sigma \in S(N). \end{aligned} \tag{1.13}$$

Here X_i is $m \times k_i$, y_i is $m \times 1$, ε_i is $m \times 1$, β_i is $k_i \times 1$, $n = Nm$, $k = \sum_{j=1}^{N} k_j$ and $\operatorname{diag}\{X_1, \dots, X_N\}$ denotes the block diagonal matrix. The (unrestricted) Zellner estimator in this model is the GLSE $\hat{\beta}_{\mathrm{ZE}} = \hat{\beta}(\hat{\Omega}_{\mathrm{ZE}})$ with $\hat{\Omega}_{\mathrm{ZE}} = S \otimes I_m$ [Revankar (1974)], where

$$S = Y'\left[I_m - X_*(X_*'X_*)^+ X_*'\right]Y. \tag{1.14}$$

Here $Y = [y_1, \dots, y_N]$ is $m \times N$, $X_* = [X_1, \dots, X_N]$ is $m \times k$ and A^+ denotes the Penrose generalized inverse of A. As is well known, the Zellner estimator $\hat{\beta}_{\mathrm{ZE}} = \hat{\beta}(\hat{\Omega}_{\mathrm{ZE}})$ satisfies (1.4). In the case of $N = 2$ (two-equation SUR model), Kariya (1981) derived the expected LUB for $\operatorname{Cov}(\hat{\beta}_{\mathrm{ZE}})$ in (1.12) as

$$\alpha_l\left(\hat{\Omega}_{\mathrm{ZE}}\right) = 1 + 2/(q-3) \quad \text{with } q = m - \operatorname{rank} X_*. \tag{1.15}$$

Also when $N = 2$, Bilodeau (1990) found a GLSE $\hat{\beta}_{\mathrm{BE}} = \hat{\beta}(\hat{\Omega}_{\mathrm{BE}})$ such that the expected LUB of $\operatorname{Cov}(\hat{\beta}_{\mathrm{BE}})$ is smaller than that of $\operatorname{Cov}(\hat{\beta}_{\mathrm{ZE}})$, that is, $\alpha_l(\hat{\Omega}_{\mathrm{BE}}) < \alpha_l(\hat{\Omega}_{\mathrm{ZE}})$. However, when $N \geq 3$, the problem of finding the expected LUB for $\operatorname{Cov}(\hat{\beta}_{\mathrm{ZE}})$ remained open since it is not a trivial generalization. In Section 3, we derive the expected LUB for $\operatorname{Cov}(\hat{\beta}_{\mathrm{ZE}})$ in a general case where $N \geq 2$. The expression of $\alpha_l(\hat{\Omega}_{\mathrm{ZE}})$ for $N \geq 3$ is rather complicated and involves zonal polynomials for $N \geq 4$. Hence a simpler but effective upper bound is also derived. In Section 4 in the heteroscedastic model (1.1) with

$$\Omega_0 = \operatorname{diag}\left\{\theta_1 I_{m_1}, \dots, \theta_N I_{m_N}\right\}, \tag{1.16}$$

the LUB and an effective upper bound for $\mathrm{Cov}(\hat{\beta}(\hat{\Omega}_0))$ are derived where $\hat{\theta}_i$ is the unbiased estimator of θ_i based on the residuals of $y_i = X_i \beta + \varepsilon_i$ with $\varepsilon_i \sim N(0, \theta_i I_{m_i})$. The case where $N = 2$ is treated by Kariya (1981) and Bilodeau (1990).

2. LUB for H and E[H]. First to derive the LUB for H in the sense of (1.11) for a given GLSE $\hat{\beta}(\hat{\Omega})$ which satisfies (1.4), let

$$\bar{X} = \Omega^{-1/2} X A^{-1/2} \quad \text{with } A = X'\Omega^{-1}X, \tag{2.1}$$

$$P = P(\hat{\Omega}, \Omega) = \Omega^{-1/2}\hat{\Omega}\Omega^{-1/2}, \tag{2.2}$$

$$\Phi(\bar{X}) = \left(\bar{X}'P^{-1}\bar{X}\right)^{-1}\bar{X}'P^{-2}\bar{X}\left(\bar{X}'P^{-1}\bar{X}\right)^{-1}, \tag{2.3}$$

$$F(n, k) = \{Z; n \times k \mid Z'Z = I_k\}. \tag{2.4}$$

Then $\bar{X} \in F(n, k)$, $\mathrm{Cov}(\hat{\beta}_{\mathrm{GME}}) = \mathrm{Cov}(\hat{\beta}(\Omega)) = A^{-1}$ and

$$H = H(\hat{\Omega}, \Omega) = \mathrm{Cov}\left(\hat{\beta}(\hat{\Omega}) \mid \hat{\Omega}\right) = A^{-1/2}\Phi(\bar{X})A^{-1/2}. \tag{2.5}$$

Therefore, by (1.11) we need to derive the least value $a_l(\hat{\Omega})$ among $a(\hat{\Omega})$s satisfying

$$\Phi(\bar{X}) \le a(\hat{\Omega})I. \tag{2.6}$$

For this purpose, observe

$$\Phi(\bar{X}\Delta) = \Delta'\Phi(\bar{X})\Delta \quad \text{for any } \bar{X} \in F(n, k) \text{ and } \Delta \in O(k), \tag{2.7}$$

where $O(k)$ is the group of $k \times k$ orthogonal matrices. Combining (2.6) with (2.7) implies that $a_l(\hat{\Omega})$ is the maximum root of $\Phi(\bar{X})$. In fact, we obtain the following important result.

LEMMA 2.1. *Let* $0 < \pi_1 \le \cdots \le \pi_n$ *be the latent roots of* P *in* (2.2). *Then* $a_l(\hat{\Omega}) = a_0(P)$ *with*

$$a_0(P) = \frac{(\pi_1 + \pi_n)^2}{4\pi_1\pi_n}. \tag{2.8}$$

That is, $a_0(P)$ *is the LUB of* $\Phi(\bar{X})$ *in* (2.6).

PROOF. For any given $\bar{X} \in F(n, k)$, choose $\Delta_1 \in O(k)$ such that $\Delta_1'\bar{X}'P^{-1}\bar{X}\Delta_1$ is diagonal. Then $Z \equiv \bar{X}\Delta_1 \in F(n, k)$ and by (2.7),

$$\Phi(Z) = (\phi_{ij}) \tag{2.9}$$

with

$$\phi_{ij} = \phi(z_i, z_j) = z_i'P^{-2}z_j/\left(z_i'P^{-1}z_i\right)\left(z_j'P^{-1}z_j\right), \tag{2.10}$$

where $Z = (z_1, \dots, z_k)$. Next choose $\Delta_2 \in O(k)$ such that $\Delta_2'\Phi(Z)\Delta_2 = \Phi(Z\Delta_2)$ is diagonal. Then setting $Z\Delta_2 = U = (u_1, \dots, u_k) \in F(n, k)$ yields

$$\Phi(U) = \mathrm{diag}\{\phi(u_1, u_1), \dots, \phi(u_k, u_k)\}, \tag{2.11}$$

since $\phi(u_i, u_j) = 0$ $(i \neq j)$. Therefore, to find the LUB a for which $\Phi(U) \leq aI$, we maximize $\phi(u_j, u_j)$ under $u_j' u_j = 1$:

$$\begin{aligned}
\sup_{u_j' u_j = 1} \phi(u_j, u_j) &= \sup_{u_j' u_j = 1} \frac{(u_j' P^{-2} u_j)(u_j' u_j)}{(u_j' P^{-1} u_j)^2} \\
&= \sup_{u_j \in R^n} \frac{(u_j' P^{-1} u_j)(u_j' P u_j)}{(u_j' u_j)^2} \\
&= \sup_{u_j' u_j = 1} (u_j' P^{-1} u_j)(u_j' P u_j) \\
&= (\pi_1 + \pi_n)^2 / 4\pi_1 \pi_n.
\end{aligned} \tag{2.12}$$

Here the last equality follows from the Kantorovich inequality [Anderson (1971), page 570]. This completes the proof. □

The LUB $a_0(P)$ in (2.8) is symmetric in π_1 and π_n and it is invariant under the transformation $P \to P^{-1}$. In fact, letting $\tilde{\pi}_1 \leq \cdots \leq \tilde{\pi}_n$ be the latent roots of P^{-1},

$$(\pi_1 + \pi_n)^2 / 4\pi_1 \pi_n = (\tilde{\pi}_1 + \tilde{\pi}_n)^2 / 4\tilde{\pi}_1 \tilde{\pi}_n. \tag{2.13}$$

Also $a_0(P) \geq 1$, where equality holds if and only if $\pi_1 = \pi_n$ or equivalently $P = \gamma I$ for some $\gamma > 0$. In other words, by (2.2), $a_0(P) = 1$ if and only if $\hat{\Omega} = \gamma \Omega$ for some $\gamma > 0$. Therefore, as a function of $P = \Omega^{-1/2} \hat{\Omega} \Omega^{-1/2}$, $a_0(P)$ is a measure of sphericity of P and hence it may be regarded as a loss function for choosing an estimator $\hat{\Omega}$ of Ω in the GLSE $\hat{\beta}(\hat{\Omega})$. This idea was adopted in Bilodeau (1990) and applied to the estimation problems in the SUR model of two equations and in the heteroscedastic model of two distinct variances.

Now by Lemma 2.1, we obtain our main result in this section.

THEOREM 2.1. *For any GLSE $\hat{\beta}(\hat{\Omega})$ satisfying* (1.4),

$$\mathrm{Cov}(\hat{\beta}_{\mathrm{GME}}) \leq \mathrm{Cov}(\hat{\beta}(\hat{\Omega})) \leq \alpha_l(\hat{\Omega}) \mathrm{Cov}(\hat{\beta}_{\mathrm{GME}}), \tag{2.14}$$

where

$$\alpha_l(\hat{\Omega}) = E[a_0(P)] = E\left[(\pi_1 + \pi_n)^2 / 4\pi_1 \pi_n\right]. \tag{2.15}$$

The expected LUB $\alpha_l(\hat{\Omega})$ may be regarded as a risk function of $\hat{\Omega}$ in choosing $\hat{\Omega}$ of $\hat{\beta}(\hat{\Omega})$ as has been discussed above.

Since the OLSE $\hat{\beta}_{\mathrm{OLSE}} = \hat{\beta}(I)$ satisfies (1.4), we obtain the following result.

COROLLARY 2.1. $\mathrm{Cov}(\hat{\beta}_{\mathrm{GME}}) \leq \mathrm{Cov}(\hat{\beta}(I)) \leq \alpha_l(I) \mathrm{Cov}(\hat{\beta}_{\mathrm{GME}})$ *with*

$$\alpha_l(I) = (\omega_1 + \omega_n)^2 / 4\omega_1 \omega_n, \tag{2.16}$$

where $\omega_1 \leq \cdots \leq \omega_n$ are the latent roots of Ω.

Though the expected LUB $\alpha_l(\hat{\Omega})$ is given by (2.15), in general it is not easy to evaluate $\alpha_l(\hat{\Omega})$ explicitly except for some special cases such as in (2.16). Further, even if it is evaluated explicitly, it may be very complicated. Hence we next derive a simpler but effective upper bound for $\mathrm{Cov}(\hat{\beta}(\hat{\Omega}))$. Such a bound is obtained through the following lemma.

LEMMA 2.2. *For $a_0(P)$ in* (2.8), $a_0(P) \le a_1(P)$, *for any $P \in S(n)$, where*

$$a_1(P) = \left\{ \frac{1}{n} \sum_{j=1}^{n} \pi_j \Big/ \prod_{j=1}^{n} \pi_j^{1/n} \right\}^n = \frac{(\mathrm{tr}\, P)^n}{n^n |P|}. \tag{2.17}$$

The equality holds if and only if $\pi_2 = \cdots = \pi_{n-1} = (\pi_1 + \pi_n)/2$.

PROOF. By the inequality between the arithmetic and geometric means,

$$\begin{aligned} a_1(P) &= \left(\frac{1}{n} \sum_{j=1}^{n} \pi_j \right)^n \Bigg/ (\pi_1 \pi_n) \left[\prod_{j=2}^{n-1} \pi_j^{1/(n-2)} \right]^{(n-2)} \\ &\ge \left(\frac{1}{n} \sum_{j=1}^{n} \pi_j \right)^n \Bigg/ (\pi_1 \pi_n) \left[\frac{1}{n-2} \sum_{j=2}^{n-1} \pi_j \right]^{(n-2)}, \end{aligned} \tag{2.18}$$

where the equality holds if and only if $\pi_2 = \cdots = \pi_{n-1} = c$ (say). Hence the smallest lower bound for $a_1(P)$ is given by

$$a_2 = \left[\frac{1}{n} (2b + (n-2)c) \right]^n \Bigg/ (\pi_1 \pi_n) c^{(n-2)},$$

where $b = (\pi_1 + \pi_n)/2$. Again using the inequality between the arithmetic and geometric means, $(1/n)(2b + (n-2)c) \ge b^{2/n} c^{(n-2)/n}$ and hence

$$a_2 \ge (b^{2/n})^n / (\pi_1 \pi_n) = b^2 / (\pi_1 \pi_n) = a_0(P),$$

where the equality holds if and only if $b = c$, proving the result. □

THEOREM 2.2.

$$\alpha_l(\hat{\Omega}) \le \alpha_1(\hat{\Omega}) = E[a_1(P)] = E\left[\left(\frac{1}{n} \sum_{j=1}^{n} \pi_j \right)^n \Bigg/ \prod_{j=1}^{n} \pi_j \right]. \tag{2.19}$$

The expectation in (2.19) is more easily evaluated as will be shown in the SUR model and a heteroscedastic model.

3. *N*-equation SUR model. In the case of the N-equation SUR model (1.13) with $\Omega = \Sigma \otimes I_m$, we first evaluate the LUB $\alpha_l(\hat{\Omega}_{\mathrm{ZE}})$ in (2.15) for $\mathrm{Cov}(\hat{\beta}_{\mathrm{ZE}})$ of the unrestricted Zellner estimator $\hat{\beta}_{\mathrm{ZE}} = \hat{\beta}(\hat{\Omega}_{\mathrm{ZE}})$ with

$$\hat{\Omega}_{\mathrm{ZE}} = S \otimes I_m \quad \text{with } S \text{ in (1.14)}. \tag{3.1}$$

As $P_{\mathrm{ZE}} = P(\hat{\Omega}_{\mathrm{ZE}}, \Omega) = \Omega^{-1/2}\hat{\Omega}_{\mathrm{ZE}}\Omega^{-1/2} = W \otimes I_m$ with $W = \Sigma^{-1/2}S\Sigma^{-1/2}$,

$$\alpha_l\big(\hat{\Omega}_{\mathrm{ZE}}\big) = E\Big[(w_1 + w_N)^2/4w_1w_N\Big], \tag{3.2}$$

where $w_1 \le \cdots \le w_N$ are the latent roots of W, where $W \sim W_N(I_N, q)$, the Wishart distribution with degrees of freedom $q = m - \operatorname{rank} X_*$. To evaluate (3.2), let $C_\kappa(\Lambda)$ with $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_N)$ denote a zonal polynominal corresponding to the partition κ of k into not more than N parts, which is denoted by $\kappa = (k_1, \dots, k_N) = \kappa(k, N)$, and let

$$\Gamma_r(x) = \pi^{r(r-1)/4}\prod_{j=1}^{r}\Gamma(x - (j-1)/2), \tag{3.3}$$

and

$$\Gamma_r(x;\kappa) = \pi^{r(r-1)/4}\prod_{j=1}^{r}\Gamma\big(x + k_j - (j-1)/2\big), \tag{3.4}$$

with $\kappa = \kappa(k, r)$. Further let $B(a, b)$ be the beta function, $(a)_r = a(a+1)\cdots(a+r-1)$, $(a)_\kappa = \prod_{j=1}^{r}(a-(j-1)/2)_{k_j}$ and

$$C_0(N,q) = \frac{\pi^{2N-2}\Gamma_{N-2}((N-2)/2)\Gamma_{N-2}((N+1)/2)}{4N^{Nq/2}\Gamma_N(N/2)\Gamma_N(q/2)}. \tag{3.5}$$

THEOREM 3.1. *Let $q > N+1$. Then the LUB $\alpha_l(\hat{\Omega}_{\mathrm{ZE}})$ is given by*

$$\begin{aligned}\alpha_l\big(\hat{\Omega}_{\mathrm{ZE}}\big) &= C_0(N,q)\\ &\quad\times\sum_{k=0}^{\infty}\frac{\Gamma(Nq/2+k)}{N^k k!}\\ &\qquad\times\sum_{\eta=0}^{2}a_\eta\sum_{r=0}^{\infty}\sum_{\rho}\Bigg\{\bigg(\frac{N-q+1}{2}\bigg)_\rho\\ &\qquad\times B\bigg(\frac{(N-1)(N+2)}{2}+k+\eta+r,\frac{q-N-1}{2}\bigg)\frac{1}{r!}\\ &\qquad\times\sum_{s=0}^{k}\binom{k}{s}\sum_{\sigma}\sum_{\tau}g_{\rho\sigma}^{\tau}\frac{\Gamma_{N-2}((N+1)/2;\tau)}{\Gamma_{N-2}(N+1;\tau)}C_\tau(I_{N-2})\Bigg\},\end{aligned} \tag{3.6}$$

where $\rho = \rho(r, N-2)$, $\sigma = \sigma(s, N-2)$, $\tau = \tau(r+s, N-2)$, and $g_{\rho\sigma}^{\tau}s$ are defined by

$$C_\rho(\Lambda)C_\sigma(\Lambda) = \sum_\tau g_{\rho\sigma}^{\tau}C_\tau(\Lambda) \tag{3.7}$$

and $(a_0, a_1, a_2) = (4, -4, 1)$.

The proof is given in the Appendix.

The expression (3.6) involves zonal polynomials and is complicated. However, in the case of the three-equation SUR model, a simpler expression is obtained as follows.

THEOREM 3.2. *When $N = 3$ and $q > 4$,*

$$\alpha_l(\hat{\Omega}_{\mathrm{ZE}}) = C_1(q) \times \sum_{k=0}^{\infty} \frac{\Gamma(3q/2+k)}{3^k k!} \times \sum_{\eta=0}^{2} \left\{ a_\eta \frac{\Gamma(k+\eta+5)}{\Gamma(q/2+k+\eta+3)} \times \sum_{s=0}^{k} \binom{k}{s} \frac{\Gamma(s+2)}{\Gamma(s+4)} \times \sum_{j=0}^{\infty} \frac{(-q/2+2)_j (k+\eta+5)_j (s+2)_j}{(q/2+k+\eta+3)_j (s+4)_j j!} \right\} \tag{3.8}$$

where $C_1(q) = \pi^{9/2}\Gamma(q/2-2)/4\ 3^{3q/2}\Gamma_3(3/2)\Gamma_3(q/2)$.

PROOF. The pdf of the roots $w = (w_1, w_2, w_3)$ is given by

$$f_3(w) = d_3(q)\exp\left\{-\tfrac{1}{2}\sum_{j=1}^{3} w_j\right\} \prod_{j=1}^{3} w_j^{\delta} \prod_{i<j} (w_j - w_i) I_A, \tag{3.9}$$

where $d_3 \equiv d_3(q) = 4C_1(q)3^{3q/2}/2^{3q/2}\Gamma(q/2-2)$, $\delta = (q-4)/2$ and I_A denotes the indicator function of the set $A = \{0 < w_1 < w_2 < w_3\}$. To evaluate $\alpha_l(\hat{\Omega}_{\mathrm{ZE}}) = \int a_l(\hat{\Omega}_{\mathrm{ZE}}) f_3(w)\, dw$, transforming w_js into $v_j = 1 - w_j/w_3$ ($j = 1, 2$) and $v_3 = w_3$ yields

$$\alpha_l(\hat{\Omega}_{\mathrm{ZE}}) = [d_3/4]\int_{B_2} \frac{(2-v_1)^2}{1-v_1}\left[\prod_{j=1}^{2}(1-v_j)^{\delta}\right](v_1-v_2)v_1v_2 \times H(v_1, v_2)\, dv_1\, dv_2, \tag{3.10}$$

where $B_2 = \{0 < v_2 < v_1 < 1\}$ and

$$\begin{aligned} H(v_1, v_2) &= \int_0^{\infty} v_3^{3q/2-1} \exp\left\{-\frac{3}{2}v_3\right\}\exp\left\{\frac{1}{2}v_3(v_1+v_2)\right\} dv_3 \\ &= \left(\frac{2}{3}\right)^{3q/2} \sum_{k=0}^{\infty} \frac{\Gamma(3q/2+k)}{3^k k!}(v_1+v_2)^k. \end{aligned} \tag{3.11}$$

Further transforming $u_1 = v_1$ and $u_2 = v_2/v_1$ in (3.10) yields

$$\alpha_l(\hat{\Omega}_{\mathrm{ZE}}) = \left(\frac{d_3}{4}\right)\left(\frac{2}{3}\right)^{3q/2} \sum_{k=0}^{\infty} \frac{\Gamma(3q/2+k)}{3^k k!} \int_0^1 (1-u_2)u_2(1+u_2)^k \times G(u_2)\,du_2, \tag{3.12}$$

where with $b(\eta, k) = a_\eta \Gamma(\delta)\Gamma(k+\eta+5)/\Gamma(\delta+5+k+\eta)$,

$$\begin{aligned} G(u_2) &= \int_0^1 (2-u_1)^2(1-u_1u_2)^\delta u_1^{4+k}(1-u_1)^{\delta-1}\,du_1 \\ &= \sum_{\eta=0}^{2} b(\eta, k) \sum_{j=0}^{\infty} \frac{(-\delta)_j(k+\eta+5)_j}{(\delta+5+k+\eta)_j j!} u_2^j. \end{aligned} \tag{3.13}$$

Hence expanding $(1+u_2)^k$ and evaluating (3.12) yields the results. In the evaluation, the monotone convergence theorem is used. □

It is noted that $\alpha_l(\hat{\Omega}_{\mathrm{ZE}})$ does not depend on unknown parameters.

The expected LUB $\alpha_l(\hat{\Omega}_{\mathrm{ZE}})$ in (3.6) or even in (3.8) is very complicated. A weaker but much simpler upper bound for $\mathrm{Cov}(\hat{\beta}_{\mathrm{ZE}})$ is obtained via Theorem 2.2 as follows.

THEOREM 3.3. *For any $N \ge 2$,*

$$\alpha_l(\hat{\Omega}_{\mathrm{ZE}}) \le \prod_{j=1}^{N-1} \left\{1 + \frac{[(N-2)/N]j + 2}{q-j-2}\right\} \equiv \gamma(N, q). \tag{3.14}$$

Moreover $\alpha_l(\hat{\Omega}_{\mathrm{ZE}}) \to 1$ as $q \to \infty$.

PROOF. By Theorem 2.2, $\alpha_1(\hat{\Omega}_{\mathrm{ZE}}) \le E\{(w_1 + \cdots + w_N)^N / N^N w_1 \cdots w_N\}$, the right-hand side of which is evaluated by Khatri and Srivastava (1971) as

$$\Gamma(Nq/2)\Gamma_N(q/2-1)/N^N\Gamma(Nq/2-N)\Gamma_N(q/2). \tag{3.15}$$

This leads to the results in Theorem 3.3. □

In the case of $N = 2$, the equality holds in (3.14), which is the result of Kariya (1981). When $N > 2$, the inequality is strict. In particular, when $N = 3$,

$$\alpha_l(\hat{\Omega}_{\mathrm{ZE}}) < \left(1 + \frac{7}{3(q-3)}\right)\left(1 + \frac{8}{3(q-4)}\right).$$

The larger the number N of equations is, the greater the upper bound $\gamma(N, q)$ in (3.14) is relative to $\alpha_l(\hat{\Omega}_{\mathrm{ZE}})$, though $\gamma(N, q)$ gives an effective bound.

For the OLSE $\hat{\beta}_{\mathrm{OLSE}} = \hat{\beta}(I)$, by Corollary 2.1,

$$\alpha_l(I) = \frac{\left(\lambda_{(1)} + \lambda_{(N)}\right)^2}{4\lambda_{(1)}\lambda_{(N)}},$$

where $\lambda_{(1)} \le \cdots \le \lambda_{(N)}$ are the ordered latent roots of Σ. Clearly, even if $q \to \infty$, $\alpha_l(I)$ does not converge to 1, implying an inefficiency of the OLSE. The LUB $\alpha_l(I)$ measures the sphericity of Σ, while $\alpha_l(\hat{\Omega}_{\mathrm{ZE}})$ measures the sphericity of $\Sigma^{-1/2}S\Sigma^{-1/2}$. As has been discussed in Section 2, $\alpha_l(\hat{\Omega})$ can be regarded as a risk function for choosing $\hat{\Omega}$ in the GLSE $\hat{\beta}(\hat{\Omega})$. When q is large, $\alpha_l(\hat{\Omega}_{\mathrm{ZE}}) < \alpha_l(I)$ unless Σ is spherical.

The efficiency of the Zellner estimator relative to the GME in terms of the generalized variance is

$$1 \le \eta = \left|\operatorname{Cov}\left(\hat{\beta}_{\mathrm{ZE}}\right)\right| / \left|\operatorname{Cov}\left(\hat{\beta}_{\mathrm{GME}}\right)\right| \le \left\{\alpha_l\left(\hat{\Omega}_{\mathrm{ZE}}\right)\right\}^k,$$

which converges to 1 as $q \to \infty$.

4. *N*-equation heteroscedastic model. In this section, we derive the expected LUB for the covariance matrix of the GLSE $\hat{\beta}(\hat{\Omega}_0)$ in the N-equation heteroscedastic model (1.1) with $\Omega = \Omega_0$ where

$$\Omega_0 = \operatorname{diag}\{\theta_1 I_{m_1}, \dots, \theta_N I_{m_N}\} \tag{4.1}$$

and

$$\hat{\Omega}_0 = \operatorname{diag}\left\{\hat{\theta}_1 I_{m_1}, \dots, \hat{\theta}_N I_{m_N}\right\}. \tag{4.2}$$

Here letting $y_i = X_i\beta + \varepsilon_i$ with $\varepsilon_i \sim N(0, \theta_i I_{m_i})$ be the ith homoscedastic submodel of (1.1), $\hat{\theta}_i$ is given by

$$\hat{\theta}_i = y_i'\left[I - X_i(X_i'X_i)^{+}X_i'\right]y_i/q_i \tag{4.3}$$

with $q_i = m_i - \operatorname{rank} X_i$, where $q_i > 0$ is assumed. Let

$$P = P\left(\hat{\Omega}_0, \Omega_0\right) = \Omega_0^{-1/2}\hat{\Omega}_0\Omega_0^{-1/2} \tag{4.4}$$

and let $v_{(1)} \le \cdots \le v_{(N)}$ be the ordered values of $v_i = \hat{\theta}_i/\theta_i$s, which are the latent roots of P. Then, by Theorem 2.1, the expected LUB for $\operatorname{Cov}(\hat{\beta}(\hat{\Omega}_0))$ is given by

$$\alpha_l\left(\hat{\Omega}_0\right) = E\left[\left(v_{(1)} + v_{(N)}\right)^2 / 4v_{(1)}v_{(N)}\right]. \tag{4.5}$$

Though the expression in (4.5) is simple, the evaluation is difficult. Therefore, we do not pursue a further evaluation beyond (4.5). On the other hand, applying Theorem 2.2 with

$$a_0(P) = \frac{\left(v_{(1)} + v_{(N)}\right)^2}{4v_{(1)}v_{(N)}} \le \frac{\left(v_{(1)} + \cdots + v_{(N)}\right)^N}{N^N v_{(1)} \cdots v_{(N)}} = a_1(P) \tag{4.6}$$

yields the following result.

THEOREM 4.1. *When* $q_i \geq 3$ $(i = 1, \ldots, N)$,

$$\alpha_l(\hat{\Omega}_0) \leq \frac{1}{N^N} \sum \binom{N}{N_1 \cdots N_N} \prod_{j=1}^{N} \left(\frac{2}{q_j}\right)^{N_j - 1} \frac{\Gamma(q_j/2 + N_j - 1)}{\Gamma(q_j/2)} \equiv \gamma_0(q_1, \ldots, q_N), \tag{4.7}$$

where $N_i \geq 0$ *and* $\sum N_j = N$. *Moreover*, $\alpha_l(\hat{\Omega}_0) \to 1$ *as all* $q_j s \to \infty$.

PROOF. From (4.6), (4.7) follows easily as

$$E[a_1(P)] = E\left[(v_1 + \cdots + v_N)^N / N^N v_1 \cdots v_N\right]$$

and as $q_j v_j \sim \chi^2(q_j)$ independently. Next let $\phi_j \equiv (2/q_j)^{N_j - 1}\Gamma(q_j/2 + N_j - 1)/\Gamma(q_j/2)$. Then $\phi_j = 1 + 2/(q_j - 2)$ if $N_j = 0$, $\phi_j = 1$ if $N_j = 1$ and $\phi_j = \prod_{i=0}^{N_j - 2}(1 + (2i)/q_j)$ if $N_j \geq 2$. Hence $\phi_j \to 1$ as $q_j \to \infty$, proving the result. □

When $N = 2$, the equality holds in (4.6) and

$$\alpha_l(\hat{\Omega}_0) = 1 + \frac{1}{2(q_1 - 2)} + \frac{1}{2(q_2 - 2)}$$

[see Kariya (1981)]. The upper bound γ_0 in Theorem 4.1 will be an effective bound for $\mathrm{Cov}(\hat{\beta}(\hat{\Omega}_0))$.

For the OLSE $\hat{\beta}_{\mathrm{OLSE}} \equiv \hat{\beta}(I)$, letting $\theta_{(1)} \leq \cdots \leq \theta_{(N)}$ be the ordered values of θ_js,

$$\alpha_l(I) = \frac{\left(\theta_{(1)} + \theta_{(N)}\right)^2}{4\theta_{(1)}\theta_{(N)}},$$

which does not converge to 1 even if q_js $\to \infty$. In fact, $\alpha_l(I)$ simply measures the sphericity of Ω_0, while $\alpha_l(\hat{\Omega}_0)$ measures the sphericity of $P_0 = \Omega_0^{-1/2}\hat{\Omega}_0\Omega_0^{-1/2}$. Since $\alpha_l(\hat{\Omega}_0) \to 1$ as q_js $\to \infty$ by Theorem 4.1, $\alpha_l(\hat{\Omega}_0) < \alpha_l(I)$ for q_js large.

APPENDIX

PROOF OF THEOREM 3.1. The joint pdf of $w = (w_1, \ldots, w_N)$ is given by

$$f(w) = d_N \exp\left\{-\tfrac{1}{2}\sum_{j=1}^{N} w_j\right\} \prod_{j=1}^{N} w_j^{\delta} \prod_{i<j} (w_j - w_i) I_A, \tag{A.1}$$

where $d_N = \pi^{N^2/2}/2^{Nq/2}\Gamma_N(N/2)\Gamma_N(q/2)$, $\delta = (q - N - 1)/2$ and $A = \{0 < w_1 < \cdots < w_N\}$. To evaluate $\alpha \equiv E[a_0(P)]$ with $a_0(P) = (w_1 +$

$w_N)^2/4w_1w_N$, transform w_is into $v_j = 1 - w_j/w_N$ ($j = 1, \dots, N-1$) and $v_N = w_N$. Then

$$\text{(A.2)}\quad \begin{aligned}\alpha = \left(\frac{d_N}{4}\right)\int_{B_{N-1}} \frac{(2-v_1)^2}{1-v_1} \prod_{j=1}^{N-1}(1-v_j)^{\delta} \prod_{i<j}^{N-1}(v_i - v_j) \prod_{j=1}^{N-1} v_j \\ \times H_{N-1}(v) \prod_{j=1}^{N-1} dv_j\end{aligned}$$

with $B_{N-1} = \{0 < v_{N-1} < \cdots < v_1 < 1\}$, $v = (v_1, \dots, v_{N-1})$ and

$$\text{(A.3)}\quad \begin{aligned} H_{N-1}(v) &= \int_0^{\infty} v_N^{Nq/2-1} \exp\left\{-\frac{N}{2}v_N\right\} \exp\left\{\frac{v_N}{2}\sum_{j=1}^{N-1} v_j\right\} dv_N \\ &= \left(\frac{2}{N}\right)^{Nq/2} \sum_{k=0}^{\infty} \phi_N(k)\left(\sum_{j=1}^{N-1} v_j\right)^k, \end{aligned}$$

where $\phi_N(k) = \Gamma(Nq/2 + k)/N^k k!$ and the monotone convergence theorem was used in the second equality. Next transforming v_js into $u_1 = v_1$, $u_j = v_j/v_1$ ($j = 2, \dots, N-1$) in (A.2) yields

$$\text{(A.4)}\quad \begin{aligned} \alpha &= (d_N/4)(2/N)^{Nq/2} \\ &\quad\times \sum_{k=0}^{\infty} \phi_N(k) \int_{B_{N-2}} \prod_{j=2}^{N-1}(1-u_j) \prod_{2\le i<j\le N-1}(u_i - u_j) \\ &\quad\times \prod_{j=2}^{N-1} u_j \left(\sum_{j=2}^{N-1} u_j + 1\right)^k G_{N-2}(u) \prod_{j=2}^{N-1} du_j \end{aligned}$$

with $B_{N-2} = \{0 < u_{N-1} < \cdots < u_2 < 1\}$, $u = (u_2, \dots, u_{N-1})$ and

$$\text{(A.5)}\quad G_{N-2}(u) = \int_0^1 (2-u_1)^2 \prod_{j=2}^{N-1}(1-u_ju_1)^{\delta} u_1^{\lambda+k-1}(1-u_1)^{\delta-1}\, du_1,$$

where $\lambda = (N-1)(N+2)/2$. Here note

$$\prod_{j=2}^{N-1}(1-u_ju_1)^{\delta} = |I_{N-2} - u_1\Lambda|^{\delta} = \sum_{r=0}^{\infty}\sum_{\rho}(-\delta)_{\rho} u_1^r C_{\rho}(\Lambda)/r!,$$

where $\Lambda = \mathrm{diag}\{u_2, \dots, u_{N-1}\}$ [see, e.g., Sugiyama (1970)]. Hence writing $(2-u_1)^2 = \sum_{\eta=0}^{2} a_{\eta} u_1^{\eta}$ and integrating (A.5) term by term yields

$$\text{(A.6)}\quad G_{N-2}(u) = \sum_{\eta=0}^{2} a_{\eta} \sum_{r=0}^{\infty}\sum_{\rho}(-\delta)_{\rho} C_{\rho}(\Lambda) B(\lambda + k + \eta + r, \delta)/r!.$$

Now to evaluate α in (A.4), expanding $(1 + \sum_{j=2}^{N-1} u_j)^k$ as

$$\sum_{s=0}^{k}\binom{k}{s}(\mathrm{tr}\,\Lambda)^s = \sum_{s=0}^{k}\binom{k}{s}\sum_{\sigma} C_{\sigma}(\Lambda)$$

and using $C_\rho(\Lambda)C_\sigma(\Lambda) = \sum_\tau g_{\rho\sigma}^\tau C_\tau(\Lambda)$,

$$\alpha = (d_N/4)(2/N)^{Nq/2}$$

$$\text{(A.7)} \qquad \times \sum_{k=0}^{\infty} \phi_N(k) \sum_{\eta=0}^{2} a_\eta \sum_{r=0}^{\infty} \sum_{\rho} (-\delta)_\rho B(\lambda + k + \eta + r, \delta)/r!$$

$$\times \sum_{s=0}^{k} \binom{k}{s} \sum_{\sigma} \sum_{\tau} g_{\rho\sigma}^{\tau} K_{N-2}(\tau)$$

with

$$K_{N-2}(\tau) = \int_{B_{N-2}} \prod_{j=2}^{N-1} (1-u_j) \prod_{i<j} (u_i - u_j) \prod_{j=2}^{N-1} u_j C_\tau(\Lambda) \prod_{j=2}^{N-1} du_j$$

$$\text{(A.8)} \qquad = \frac{\Gamma_{N-2}((N-2)/2)\ \Gamma_{N-2}((N+1)/2;\tau)\ \Gamma_{N-2}((N+1)/2)}{\pi^{(N-2)^2/2}\ \Gamma_{N-2}(N+1;\tau)}$$

$$\times C_\tau(I_{N-2}),$$

where Lemma 3.3 of Sugiyama (1966) is used and $\Gamma_N(a;\tau)$ is given by (3.4). This completes the proof. □

REFERENCES

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

BILODEAU, M. (1990). On the choice of the loss function in covariance estimation. *Statist. Decisions* **8** 131–139.

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

KARIYA, T. and TOYOOKA, Y. (1985). Nonlinear versions of the Gauss–Markov theorem and GLSE. In *Multivariate Analysis* **6** 345–354. North-Holland, Amsterdam.

KHATRI, C. G. and SRIVASTAVA, M. S. (1971). On exact non-null distributions of likelihood ratio criteria for sphericity test and equality of two covariance matrices. *Sankhyā Ser. A* **33** 201–206.

REVANKAR, N. S. (1974). Some finite sample results in the context of two seemingly unrelated regression equations. *J. Amer. Statist. Assoc.* **69** 187–190.

SUGIYAMA, T. (1966). On the distribution of the largest latent root and the corresponding latent vector for principal component analysis. *Ann. Math. Statist.* **37** 995–1001.

SUGIYAMA, T. (1970). Joint distribution of the extreme roots of a covariance matrix. *Ann. Math. Statist.* **41** 655–657.

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

ZELLNER, A. (1962). An efficient method of estimating seemingly unrelated regression and tests for aggregation bias. *J. Amer. Statist. Assoc.* **57** 348–368.

ZELLNER, A. (1963). Estimators for seemingly unrelated regressions: some exact finite sample results. *J. Amer. Statist. Assoc.* **58** 977–992.

DEPARTMENT OF ECONOMICS
YAMAGUCHI UNIVERSITY
YAMAGUCHI-SHI, YAMAGUCHI-KEN
JAPAN

INSTITUTE OF ECONOMIC RESEARCH
HITOTSUBASHI UNIVERSITY
KUNITACHI, TOKYO, 186
JAPAN
E-MAIL: cr00055@srv.cc.hit-u.ac.jp