Journal of Multivariate Analysis **83**, 37–55 (2002)
doi:10.1006/jmva.2001.2050

A Maximal Extension of the Gauss–Markov Theorem and Its Nonlinear Version[1]

Takeaki Kariya

Kyoto University, Kyoto, Japan
E-mail: kariya@kier.kyoto-u.ac.jp

and

Hiroshi Kurata[2]

University of Tokyo, Tokyo, Japan
E-mail: kurata@waka.c.u-tokyo.ac.jp

Received January 4, 1999; published online February 28, 2002

In this paper, first we make a maximal extension of the well-known Gauss–Markov Theorem (GMT) in its linear framework. In particular, the maximal class of distributions of error term for which the GMT holds is derived. Second, we establish a nonlinear version of the maximal GMT and describe some interesting families of distributions of error term for which the nonlinear GMT holds. © 2002 Elsevier Science (USA)

AMS 2000 subject classifications: 62J05; 62H12.

Key words and phrases: Gauss–Markov theorem; nonlinear versions of Gauss–Markov theorem; location-equivariant estimator; generalized least squares estimator; elliptically symmetric distribution.

1. INTRODUCTION

In this paper we aim to strengthen the Gauss–Markov Theorem (GMT). As is well known, the GMT states that when the error term ε in a linear regression model

$$y = X\beta + \varepsilon \tag{1.1}$$

[1] This research was financially supported by *The Japan Securities Scholarship Foundation*.

[2] To whom correspondence should be addressed. Portions of Kurata's work were financially supported by the Faculty of Economics Yamaguchi University and the Ministry of Education in Japan (Kiban (c)(2) No. 11630028).

0047-259X/02 $35.00

satisfies the condition

$$\mathscr{L}(\varepsilon) \in \mathscr{P}(0, \Sigma) \qquad \text{with} \quad \Sigma \in \mathscr{S}(n), \tag{1.2}$$

the Gauss–Markov estimator (GME) defined by

$$b(\Sigma) = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}y \tag{1.3}$$

minimizes the risk matrix

$$R_P(\hat{\beta}, \beta) = E_P\{(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\} \tag{1.4}$$

with respect to the ordering of nonnegative definiteness in the class of linear unbiased estimators:

$$\mathscr{C}_0 = \{\hat{\beta} \in \mathscr{C}_{\text{all}} \mid \hat{\beta} = Cy, C \text{ is a } k \times n \text{ matrix satisfying } CX = I_k\}. \tag{1.5}$$

Here X is an $n \times k$ known matrix of full rank, $P \equiv \mathscr{L}(\varepsilon)$ denotes the distribution of the error term ε, $\mathscr{C}_{\text{all}}$ the set of estimators of β, $\mathscr{S}(n)$ the set of $n \times n$ positive definite matrices, and $\mathscr{P}(0, \Sigma)$ denotes the class of distributions wit mean 0 and covariance matrix $\gamma\Sigma$ for some $\gamma > 0$. More generally, for given $\mu \in R^n$ and $\Phi \in \mathscr{S}(n)$, we set

$$\begin{aligned}\mathscr{P}(\mu, \Phi) = {} & \text{the class of distributions on } R^n \text{ with mean } \mu \text{ and} \\ & \text{finite covariance matrix } \gamma\Phi, \text{ where } \gamma > 0 \text{ is unspecified.}\end{aligned}$$

Of course, the risk matrix in (1.4) is the covariance matrix of $\hat{\beta} \in \mathscr{C}_0$ under (1.2). We shall call $b(\Sigma)$ the GME even when Σ is unknown so long as (1.2) holds. This GMT implies that, to define a measure of efficiency of an estimator in $\mathscr{C}_0$ in terms of risk matrix, the GME is the estimator to be compared with, because the covariance matrix of the GME gives the lower bound for the risk (1.4), provided $P = \mathscr{L}(\varepsilon)$ belongs to the class $\mathscr{P}(0, \Sigma)$. For example, an efficiency of an estimator in $\mathscr{C}_0$ is often defined by such a measure as $|R_P(b(\Sigma), \beta)|/|R_P(\hat{\beta}, \beta)|$ or $\text{tr}\{R_P(\hat{\beta}, \beta)\, R_P(b(\Sigma), \beta)^{-1}\}$, where $P \in \mathscr{P}(0, \Sigma)$.

However the optimality of the GME depends on the class $\mathscr{P}(0, \Sigma)$ of distributions in (1.2) as well as the class $\mathscr{C}_0$ in (1.5) of estimators. In fact, the class $\mathscr{C}_0$ is implied by unbiasedness when $\mathscr{P}(0, \Sigma)$ is first fixed. But as will be shown in the first part of this paper, this is not true when $\mathscr{P}(0, \Sigma)$ is maximally broadened, say to $\mathscr{P}_{\max}(\Sigma)$, for the GMT to hold with $\mathscr{C}_0$ first fixed. Notably $\mathscr{P}_{\max}(\Sigma)$ contains some distributions in $\mathscr{P}(\mu, \Phi)$ where $\mu \neq 0$ and $\Phi \neq \Sigma$. In Section 2, the problem is formally set up, and the optimality of an estimator is defined relative to both a class of estimators and a class of error distributions. In Sections 3 and 4, based on the definition and the equivalence relation introduced on the set $\mathscr{S}(n)$ of positive definite matrices, a maximal extension of the GMT is made in its linear framework.

In application, the model in (1.1) and (1.2) is often specified in such a way that Σ is unknown but is a function of parameters θ, say $\Sigma = \Sigma(\theta)$, which is estimable by the residuals of the ordinary least squares estimator (OLSE). Typically such models appear as errors are heteroscedastic or autoregressive. In such a case, $b(\hat{\Sigma})$ is often used with Σ in (1.3) replaced by an estimator $\hat{\Sigma}$. We call such an estimator a generalized least squares estimator (GLSE). In Section 5, a nonlinear version of the GMT is established. First it is shown that when $P(\varepsilon = 0) = 0$, the class $\mathscr{C}_1$ of location equivariant estimators is equivalent to the class of GLSE. Second, the GME is shown to be optimal relative to a certain pair $(\mathscr{D}_1, \mathscr{Q}_1(\Sigma))$ with $\mathscr{D}_1 \subset \mathscr{C}_1$ and $\mathscr{Q}_1(\Sigma) \subset \mathscr{P}_{\max}(\Sigma)$. The result is an extension of Kariya (1985), Kariya and Toyooka (1985), and Eaton (1985).

In the literature, adopting concentration probability as a criterion for comparison of estimators, Berk and Hwang (1989), which can be viewed as an extension of Hwang (1985), Kuritsyn (1986), and Ali and Ponnapalli (1990), established some linear or nonlinear extensions of the GMT. Some related results are found, for example, in Andrews and Phillips (1987), Eaton (1986, 1988), and Jensen (1996). In particular, in Eaton (1988), a class of distributions that possesses a certain kind of symmetric property under which the GME is optimal in terms of concentration probability is described.

2. FORMULATION OF THE PROBLEM

To state our problem more specifically, let $\mathscr{P}_{\text{all}}$ denote the set of all distributions on R^n with second moments. The optimality we use is defined by the following:

DEFINITION 2.1. For a class $\mathscr{C}(\subset \mathscr{C}_{\text{all}})$ of estimators and for a class $\mathscr{P}(\subset \mathscr{P}_{\text{all}})$ of distributions of ε, an estimator $\hat{\beta}^*$ is said to be $(\mathscr{C}, \mathscr{P})$-optimal, if $\hat{\beta}^* \in \mathscr{C}$ and

$$R_P(\hat{\beta}^*, \beta) \leqslant R_P(\hat{\beta}, \beta) \qquad \text{holds for all} \quad \hat{\beta} \in \mathscr{C} \qquad \text{and} \qquad P \in \mathscr{P}, \tag{2.1}$$

where the inequality for matrices should be understood in terms of nonnegative definiteness.

In case of the GMT, the statement along this definition becomes

$$b(\Sigma) \text{ is } (\mathscr{C}_0, \mathscr{P}(0, \Sigma))\text{-optimal.} \tag{2.2}$$

Based on Definition 2.1, we strengthen this GMT as follows. First, in Section 4, for given $\Sigma \in \mathcal{S}(n)$ and for the class $\mathcal{C}_0$, we will derive a maximal class $\mathcal{P}_{\max}(\Sigma)$ of distributions of ε for which $b(\Sigma)$ is $(\mathcal{C}_0, \mathcal{P}_{\max}(\Sigma))$-optimal, where $\mathcal{P}_{\max}(\Sigma)$ contains a class $\mathcal{P}(\mu, \Phi)$ with some $\mu \neq 0$ and $\Phi \neq \Sigma$ as well as the class $\mathcal{P}(0, \Sigma)$ for which the original GMT holds. The GME $b(\Sigma)$ is in particular $(\mathcal{C}_0, \mathcal{P}(\mu, \Phi))$-optimal if $\mathcal{P}(\mu, \Phi) \subset \mathcal{P}_{\max}(\Sigma)$. And then $\mathcal{C}_0$ is no longer the class of linear unbiased estimators under $P \in \mathcal{P}(\mu, \Phi)$, though it is a class of linear estimators. In fact, the optimality of the GME is extended even to such a class of distributions under which the GME itself is biased.

Second, in Section 5, we will extend the class $\mathcal{C}_0$ of linear estimators to the following class of nonlinear estimators,

$$\mathcal{C}_1 = \{\hat{\beta} = \hat{\beta}(y) \in \mathcal{C}_{\text{all}} \mid \hat{\beta}(y) = b(I_n) + d(e), d \text{ is a } k \times 1 \text{ vector-valued measurable function on } R^n\}, \tag{2.3}$$

and derive the maximal classes $\mathcal{D}_1 (\subset \mathcal{C}_1)$ and $\mathcal{Q}_1(\Sigma) (\subset \mathcal{P}_{\max}(\Sigma))$ for which

$$b(\Sigma) \text{ is } (\mathcal{D}_1, \mathcal{Q}_1(\Sigma))\text{-optimal},$$

where $b(I_n) = (X'X)^{-1} X'y$ is the OLSE and e is the OLS residual vector defined by

$$e = y - Xb(I_n) = Ny \qquad \text{with} \quad N = I_n - X(X'X)^{-1} X'. \tag{2.4}$$

Here, $\mathcal{D}_1$ is the class of estimators in $\mathcal{C}_1$ that have the second moments, and $\mathcal{Q}_1(\Sigma)$ is the maximal class of distributions for which (2.1) holds with $\mathcal{C} = \mathcal{D}_1$ and $\mathcal{P} = \mathcal{Q}_1(\Sigma)$. The class $\mathcal{C}_1$ in (2.3) is the class of location equivariant estimators satisfying

$$\hat{\beta}(y + Xa) = \hat{\beta}(y) + a \qquad \text{for any} \quad a \in R^k,$$

which was used in Berk and Hwang (1989) in a different context. The class $\mathcal{C}_1$ is also shown to be equivalent to the following class of nonlinear estimators

$$\tilde{\mathcal{C}}_{\text{KT}} = \{\tilde{\beta} \in \mathcal{C}_{\text{all}} \mid \tilde{\beta} = \hat{\beta} + a\chi_{\{e=0\}}, \hat{\beta} \in \mathcal{C}_{\text{KT}}, a \in R^k\}, \tag{2.5}$$

where $\chi_{\{e=0\}}$ is the indicator function of the set $\{e = 0\} \subset R^n$, and the class $\mathcal{C}_{\text{KT}}$ is defined by

$$\mathcal{C}_{\text{KT}} = \{\hat{\beta} \in \mathcal{C}_{all} \mid \hat{\beta} = C(e)\, y, C(\cdot) \text{ is a } k \times n \text{ matrix-valued measurable function on } R^n \text{ which satisfies } C(\cdot)\, X = I_k\}, \tag{2.6}$$

which was adopted in Kariya (1985), Kariya and Toyooka (1985), and Eaton (1985) to establish a nonlinear version of the GMT. The class $\mathscr{C}_{\mathrm{KT}}$ is a nonlinear extension of the class $\mathscr{C}_0$ in (1.5), and is directly understandable. Clearly $\mathscr{C}_0 \subset \mathscr{C}_{\mathrm{KT}} \subset \tilde{C}_{\mathrm{KT}} = \mathscr{C}_1$. The class $\mathscr{C}_{\mathrm{KT}}$ contains a GLSE of the form

$$b(\hat{\Sigma}) = (X'\hat{\Sigma}^{-1}X)^{-1}\, X'\hat{\Sigma}^{-1}y \qquad \text{with} \quad \hat{\Sigma} = \hat{\Sigma}(e), \tag{2.7}$$

which is often used when Σ is estimable by the OLS residual e. Also in Section 5, it is shown that $\mathscr{Q}_1(\Sigma)$ contains several interesting classes of distributions such as a class of elliptically symmetric distributions under an appropriate condition.

3. AN EQUIVALENCE RELATION ON $\mathscr{S}(n)$

We introduce an equivalent relation $\sim$ on the set $\mathscr{S}(n)$ of $n \times n$ positive definite matrices. Let $P \equiv \mathscr{L}(\varepsilon) \in \mathscr{P}(\mu, \Phi)$.

DEFINITION 3.1. For any Σ and Ψ in $\mathscr{S}(n)$,

$$\Sigma \sim \Psi \qquad \text{if and only if} \quad b(\Sigma) = b(\Psi) \qquad \text{a.s. } P. \tag{3.1}$$

Note that the identity $b(\Sigma) = b(\Psi)$ a.s. P is slightly different from the functional identity $B(\Sigma) = B(\Psi)$ with

$$B(\Omega) = (X'\Omega^{-1}X)^{-1}\, X'\Omega^{-1},$$

since the latter requires $b(\Sigma) = b(\Psi)$ for all $y \in R^n$ even if the distribution of y is degenerate.

Now fix $\Sigma \in \mathscr{S}(n)$ and let us describe the set of matrices equivalent to Σ: $\{\Psi \in \mathscr{S}(n) \mid \Psi \sim \Sigma\}$. To do so, let

$$\bar{X}_\Sigma = \Sigma^{-1/2}X(X'\Sigma^{-1}X)^{-1/2} \qquad \text{and} \qquad \bar{Z}_\Sigma = \Sigma^{1/2}Z(Z'\Sigma Z)^{-1/2} \tag{3.2}$$

and form the $n \times n$ orthogonal matrix

$$\Gamma_\Sigma = \begin{pmatrix} \bar{X}'_\Sigma \\ \bar{Z}'_\Sigma \end{pmatrix}, \tag{3.3}$$

where Z is an $n \times (n-k)$ matrix such that

$$ZZ' = N \qquad \text{and} \qquad Z'Z = I_{n-k},$$

and it is fixed throughout this paper. Further let $\eta = \Sigma^{-1/2}\varepsilon$ and define the two vectors $\tilde{\eta}_1$: $k \times 1$ and $\tilde{\eta}_2$: $(n-k) \times 1$ by

$$\tilde{\eta} = \Gamma_\Sigma \eta = \begin{pmatrix} \bar{X}'_\Sigma \eta \\ \bar{Z}'_\Sigma \eta \end{pmatrix} = \begin{pmatrix} \tilde{\eta}_1 \\ \tilde{\eta}_2 \end{pmatrix}. \tag{3.4}$$

Then it is easy to see that for $\Sigma \in \mathscr{S}(n)$ fixed,

$$b(\Sigma) - \beta = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}\varepsilon = (X'\Sigma^{-1}X)^{-1/2}\,\tilde{\eta}_1, \tag{3.5}$$

and for any $\Psi \in \mathscr{S}(n)$,

$$\begin{aligned} b(\Psi) - \beta &= (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}\{X(X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} + \Sigma Z(Z'\Sigma Z)^{-1} Z'\}\,\varepsilon \\ &= (X'\Sigma^{-1}X)^{-1/2}\,\tilde{\eta}_1 + (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}\Sigma Z(Z'\Sigma Z)^{-1/2}\,\tilde{\eta}_2 \\ &= \{b(\Sigma) - \beta\} + \{b(\Psi) - b(\Sigma)\}, \end{aligned} \tag{3.6}$$

where the matrix identity

$$X(X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} + \Sigma Z(Z'\Sigma Z)^{-1} Z' = I_n \tag{3.7}$$

is used in the first line of (3.6). The following theorem characterizes the equivalence relation in (3.1).

Theorem 3.1. (1) *For Σ and Ψ in $\mathscr{S}(n)$,* (i) *the relation $\Psi \sim \Sigma$ holds if and only if* (ii) *$X'\Psi^{-1}\Sigma Z = 0$, which is equivalent to* (iii) *$\Psi \in \mathscr{R}(\Sigma)$, where*

$$\begin{aligned} \mathscr{R}(\Sigma) &= \{\Phi \in \mathscr{S}(n) \mid \Phi = X\Upsilon X' + \Sigma Z\Delta Z'\Sigma,\ \Upsilon \in \mathscr{S}(k),\ \Delta \in \mathscr{S}(n-k)\} \\ &= \{\Phi \in \mathscr{S}(n) \mid \Phi^{-1} = \Sigma^{-1}X\bar{\Upsilon}X'\Sigma^{-1} + Z\bar{\Delta}Z',\ \bar{\Upsilon} \in \mathscr{S}(k),\ \bar{\Delta} \in \mathscr{S}(n-k)\}. \end{aligned} \tag{3.8}$$

(2) *$\mathscr{C}_0$ is isomorphic with $\mathscr{S}(n)/\sim$. The map*

$$\tau : \mathscr{C}_0 \to \mathscr{S}(n)/\sim, \qquad \hat{\beta} = Cy \mapsto \mathscr{R}([C'C + N]^{-1})$$

is isomorphic, where $\mathscr{R}(\Psi)$ with $\Psi^{-1} = C'C + ZZ'$ is expressed as

$$\mathscr{R}(\Psi) = \{\Omega \in \mathscr{S}(n) \mid \Omega^{-1} = C'\Upsilon C + Z\Delta Z',\ \Upsilon \in \mathscr{S}(k),\ \Delta \in \mathscr{S}(n-k)\}.$$

Proof. From (3.5) and (3.6), the equality $b(\Psi) = b(\Sigma)$ a.s. holds under P if and only if

$$E_P\{dd'\} = 0 \qquad \text{with} \quad d \equiv (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}\Sigma Z(Z'\Sigma Z)^{-1/2}\,\tilde{\eta}_2. \tag{3.9}$$

By setting $E_P(\varepsilon) = \mu$ and $Cov_P(\varepsilon) = \gamma\Phi$ with $\gamma > 0$, we obtain

$$E_P\{dd'\} = (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}\Sigma Z(Z'\Sigma Z)^{-1} Z'(\mu\mu' + \gamma\Phi) \times Z(Z'\Sigma Z)^{-1} Z'\Sigma\Psi^{-1}X(X'\Psi^{-1}X)^{-1}.$$

Hence it is shown that a necessary and sufficient condition for (3.9) is that $X'\Psi^{-1}\Sigma Z = 0$. Thus the equivalence between (i) and (ii) follows. The equivalence between (ii) and (iii) is obtained from Corollary 1 and 2 in p. 160 of Rao and Mitra (1971).

(2) Since any $\hat{\beta} = Cy \in \mathscr{C}_0$ is a GLSE $b(\Psi)$ with $\Psi^{-1} = C'C + N$ (see Theorem 3.2 of Kariya and Toyooka (1985)), the result follows from (1). ∎

By this theorem, the quotient space $\mathscr{S}(n)/\sim$ is identified with the set of equivalence classes $\{\mathscr{R}(\Omega) \mid \Omega \in \mathscr{S}(n)\}$, and $\mathscr{C}_0$ is in one-to-one correspondence with $\mathscr{B}_0 = \{b(\Omega) \mid \Omega \in \mathscr{S}(n)\}$.

4. A MAXIMAL EXTENSION OF THE GAUSS–MARKOV THEOREM

In the model (1.1), we fix a matrix $\Sigma \in \mathscr{S}(n)$ and derive the maximal class $\mathscr{P}_{\max}(\Sigma)$ of distributions of ε for which $b(\Sigma)$ is $(\mathscr{C}_0, \mathscr{P}_{\max}(\Sigma))$-optimal. Note that $\hat{\beta} = Cy$ in $\mathscr{C}_0$ is decomposed as

$$\begin{aligned} \hat{\beta} - \beta &= C\{X(X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} + \Sigma Z(Z'\Sigma Z)^{-1} Z'\}\,\varepsilon \\ &= (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}\varepsilon + C\Sigma Z(Z'\Sigma Z)^{-1} Z'\varepsilon \\ &= A^{-1/2}\tilde{\eta}_1 + H\tilde{\eta}_2, \end{aligned} \tag{4.1}$$

where $\tilde{\eta} = (\tilde{\eta}_1', \tilde{\eta}_2')'$ is defined in (3.4),

$$A = X'\Sigma^{-1}X, \qquad H = C\Sigma Z(Z'\Sigma Z)^{-1/2} \tag{4.2}$$

and the matrix identity (3.7) is used in the first line of (4.1). Then the risk matrix of $\hat{\beta}$ is expressed as

$$\begin{aligned} R_P(\hat{\beta}, \beta) &= E_P\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\} \\ &= A^{-1/2}E_P\{\tilde{\eta}_1\tilde{\eta}_1'\}\, A^{-1/2} + HE_P\{\tilde{\eta}_2\tilde{\eta}_2'\}\, H' + A^{-1/2}E_P\{\tilde{\eta}_1\tilde{\eta}_2'\}\, H' \\ &\quad + HE_P\{\tilde{\eta}_2\tilde{\eta}_1'\}\, A^{-1/2} \\ &= V_{11} + V_{22} + V_{12} + V_{21} \qquad \text{(say)}, \end{aligned} \tag{4.3}$$

where $V_{12} = V_{21}'$, V_{ij}'s depend on P and $R_P(b(\Sigma), \beta) = V_{11}$.

THEOREM 4.1. *Suppose $P \in \mathscr{P}(\mu, \Phi)$ and fix $\Sigma \in \mathscr{S}(n)$.*

(1) *A necessary and sufficient condition for $b(\Sigma)$ to be $(\mathscr{C}_0, \mathscr{P}(\mu, \Phi))$-optimal is that* (i) $E_P\{\tilde{\eta}_1 \tilde{\eta}_2'\} = 0$ *holds for any $P \in \mathscr{P}(\mu, \Phi)$, which is equivalent to the condition that* (ii) $(\mu, \Phi) \in \mathscr{M}(\Sigma) \times \mathscr{R}(\Sigma)$, *where*

$$\mathscr{M}(\Sigma) = L(X) \cup L(\Sigma Z), \tag{4.4}$$

$L(\cdot)$ denotes the linear subspace spanned by the column vectors of matrix $\cdot$ and $\mathscr{R}(\Sigma)$ is defined in (3.8).

(2) *When $b(\Sigma)$ is $(\mathscr{C}_0, \mathscr{P}(\mu, \Phi))$-optimal and $Cov_P(\varepsilon) = \gamma\Phi$ with $\Phi = X\Upsilon X' + \Sigma Z \Delta Z' \Sigma$, then the minimum risk $R_P(b(\Sigma), \beta)$ is evaluated as*

$$R_P(b(\Sigma), \beta) = \begin{cases} cc' + \gamma\Upsilon & \text{when} \quad \mu = Xc \quad \text{for some } c \in R^k \\ \gamma\Upsilon & \text{when} \quad \mu = \Sigma Z d \quad \text{for some } d \in R^{n-k}. \end{cases} \tag{4.5}$$

Proof. (1) Suppose first that (i) holds. Then for any $\hat{\beta} \in \mathscr{C}_0$ and for any $P \in \mathscr{P}(\mu, \Phi)$, the risk matrix in (4.3) is expressed as $R_P(\hat{\beta}, \beta) = V_{11} + V_{22}$, which is greater than $V_{11} = R_P(b(\Sigma), \beta)$, proving the sufficiency of (i). Conversely, suppose that $b(\Sigma)$ is $(\mathscr{C}_0, \mathscr{P}(\mu, \Phi))$-optimal. Since for any $a \in R$ and any $F\colon (n-k) \times k$, the estimator of the form $\hat{\beta} = Cy$ with $C = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} + aF'Z'$ belongs to $\mathscr{C}_0$, the risk matrix is bounded below by that of $b(\Sigma)$, that is, $R_P(\hat{\beta}, \beta) \geqslant R_P(b(\Sigma), \beta)$ for any $a \in R$, $F\colon (n-k) \times k$ and $P \in \mathscr{P}(\mu, \Phi)$. For $c \in R^k$, we set

$$c' R_P(\hat{\beta}, \beta)\, c = a^2 f_2 + 2a f_1 + f_0,$$

where f_i values are defined as

$$\begin{aligned} f_0 &= c'V_{11}c = c'R_P(b(\Sigma), \beta)\, c \\ &\qquad \text{with} \quad V_{11} = A^{-1/2} E_P\{\tilde{\eta}_1 \tilde{\eta}_1'\}\, A^{-1/2}, \\ f_1 &= c'\{V_{12} + V_{21}\}\, c \\ &\qquad \text{with} \quad V_{12} = V_{21}' = A^{-1/2} E_P\{\tilde{\eta}_1 \tilde{\eta}_2'\} (Z'\Sigma Z)^{1/2} F, \\ f_2 &= c'V_{22}c \\ &\qquad \text{with} \quad V_{22} = F'(Z'\Sigma Z)^{1/2}\, E_P\{\tilde{\eta}_2 \tilde{\eta}_2'\} (Z'\Sigma Z)^{1/2} F. \end{aligned} \tag{4.6}$$

Note that f_i values are free from a, although they depend on c, F, and P. Here it is easy to see that for some c, F, and P, the term $a^2 f_2 + 2a f_1$ can be made negative by taking $|a|$ sufficiently small, unless $f_1 = 0$ for any c, F, and P, or equivalently, $V_{12} + V_{21} = V_{12} + V'_{12} = 0$ for any F and P. From this, (i) follows.

Next to show that (i) and (ii) are equivalent, let $E_P(\varepsilon) = \mu$, and $Cov_P(\varepsilon) = \gamma\Phi$ with $\gamma > 0$. Then $E_P\{\tilde{\eta}_1\tilde{\eta}_2'\}$ is directly calculated as

$$E_P\{\tilde{\eta}_1\tilde{\eta}_2'\} = A^{-1/2}X'\Sigma^{-1}[\mu\mu' + \gamma\Phi]\, Z(Z'\Sigma Z)^{-1/2},$$

and the condition (i) holds if and only if

$$A^{-1/2}X'\Sigma^{-1}[\mu\mu' + \gamma\Phi]\, Z(Z'\Sigma Z)^{-1/2} = 0 \qquad \text{for any} \quad \gamma > 0,$$

which is in turn equivalent to

$$-\gamma X'\Sigma^{-1}\Phi Z = X'\Sigma^{-1}\mu\mu' Z \qquad \text{for any} \quad \gamma > 0. \tag{4.7}$$

Since the right-hand side of (4.7) does not depend on $\gamma > 0$, it is equivalent to

$$X'\Sigma^{-1}\Phi Z = 0 \qquad \text{and} \qquad X'\Sigma^{-1}\mu\mu' Z = 0.$$

By Theorem 3.1, $X'\Sigma^{-1}\Phi Z = 0$ is equivalent to $\Phi \in \mathscr{R}(\Sigma)$. On the other hand, $X'\Sigma^{-1}\mu\mu' Z = 0$ holds if and only if $X'\Sigma^{-1}\mu = 0$ or $Z'\mu = 0$, proving equivalence between (i) and (ii). Thus the proof is complete, since (2) is straightforward. ∎

COROLLARY 4.2. *For fixed $\Sigma \in \mathscr{S}(n)$, let*

$$\mathscr{P}_{\max}(\Sigma) = \bigcup \{\mathscr{P}(\mu, \Phi) \mid (\mu, \Phi) \in \mathscr{M}(\Sigma) \times \mathscr{R}(\Sigma)\}. \tag{4.8}$$

Then $\mathscr{P}_{\max}(\Sigma)$ is the maximal class of distributions for which

$$b(\Sigma) \text{ is } (\mathscr{C}_0, \mathscr{P}_{\max}(\Sigma))\text{-optimal.} \tag{4.9}$$

From the proof of Theorem 4.1, it is clear that when γ is fixed in $Cov_P(\varepsilon) = \gamma\Phi$ in the definition of $\mathscr{P}(\mu, \Phi)$, $\mu\mu' + \gamma\Phi \in \mathscr{R}(\Sigma)$ is necessary and sufficient for $E_P\{\tilde{\eta}_1\tilde{\eta}_2'\} = 0$. It is noted that $b(\lambda\Phi) = b(\Phi)$ for any $\lambda > 0$.

COROLLARY 4.3. (1) *When $b(\Phi) = b(\Sigma)$ a.s., $\mu \in \mathscr{M}(\Sigma)$ is necessary and sufficient for $b(\Sigma)$ to be $(\mathscr{C}_0, \mathscr{P}(\mu, \Phi))$-optimal.*

(2) *When $\mu \in \mathscr{M}(\Sigma)$, $\Phi \in \mathscr{R}(\Sigma)$ is necessary and sufficient for $b(\Sigma)$ to be $(\mathscr{C}_0, \mathscr{P}(\mu, \Phi))$-optimal.*

Proof. Since $b(\Phi) = b(\Sigma)$ a.s. implies $\Phi \in \mathscr{R}(\Sigma)$ under $P \in \mathscr{P}(\mu, \Phi)$, the result follows from the proof of Theorem 4.1. (2) is also clear from Theorem 4.1. ∎

EXAMPLE 4.4. Let $\mu = \Sigma Z d$ and $\Phi = X\Upsilon X' + \Sigma Z \Delta Z' \Sigma \in \mathscr{R}(\Sigma)$, where $d \in R^{n-k}$, $\Upsilon \in \mathscr{S}(k)$ and $\Delta \in \mathscr{S}(n-k)$. Suppose $P = \mathscr{L}(\varepsilon) \in \mathscr{P}(\mu, \Phi)$ with $Cov_P(\varepsilon) = \gamma\Phi$. Then $b(\Sigma) = b(\Phi)$ and $b(\Sigma)$ is unbiased, since

$$E_P\{b(\Sigma)\} = \beta + (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}\Sigma Z d = \beta.$$

The risk matrix of $b(\Sigma)$ is

$$R_P(b(\Sigma), \beta) = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}[\mu\mu' + \gamma\Phi]\, \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} = \gamma\Upsilon.$$

Further for $\hat{\beta} = Cy = \beta + C\varepsilon$ in $\mathscr{C}_0$,

$$E_P\{\hat{\beta}\} = \beta + C\Sigma Z d.$$

Hence $\mathscr{C}_0$ contains some biased estimators in general. However

$$\begin{aligned} R_P(\hat{\beta}, \beta) &= C E_P\{\varepsilon\varepsilon'\}\, C' = C[\mu\mu' + \gamma\Phi]\, C' \\ &= C\Sigma Z d d' Z' \Sigma C' + \gamma\Upsilon + C\Sigma Z \Delta Z' \Sigma C' \geqslant \gamma\Upsilon. \end{aligned}$$

Therefore $b(\Sigma)$ minimizes $R_P(\hat{\beta}, \beta)$ with respect to $\hat{\beta} \in \mathscr{C}_0$ and the minimum is $\gamma\Upsilon$.

EXAMPLE 4.5. Let $\mu = Xc$ and $\Phi = X\Upsilon X' + \Sigma Z \Delta Z' \Sigma \in \mathscr{R}(\Sigma)$, and suppose $\mathscr{L}(\varepsilon) \in \mathscr{P}(\mu, \Phi)$ with $Cov_P(\varepsilon) = \gamma\Phi$, where $c \in R^k$, $\Upsilon \in \mathscr{S}(k)$ and $\Delta \in \mathscr{S}(n-k)$. Then $b(\Sigma) = b(\Phi)$, but the GME $b(\Sigma)$ is biased:

$$E_P\{b(\Sigma)\} = \beta + (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} X c = \beta + c.$$

Then risk matrix of $b(\Sigma)$ is

$$R_P(b(\Sigma), \beta) = cc' + \gamma\Upsilon.$$

Further for $\hat{\beta} = Cy$ in $\mathscr{C}_0$,

$$R_P(\hat{\beta}, \beta) = C[\mu\mu' + \gamma\Phi]\, C' = cc' + \gamma\Upsilon + C\Sigma Z \Delta Z' \Sigma C' \geqslant cc' + \gamma\Upsilon.$$

5. NONLINEAR VERSIONS OF THE GAUSS–MARKOV THEOREM

In this section, we enlarge the class $\mathscr{C}_0$ of linear estimators to a subclass $\mathscr{D}$ of the class of location equivariant estimators $\mathscr{C}_1$ in (2.3), and establish a nonlinear version of the GMT that $b(\Sigma)$ is $(\mathscr{D}, \mathscr{Q}(\Sigma))$-optimal for some class $\mathscr{Q}(\Sigma)$ of distributions of ε, where $\Sigma \in \mathscr{S}(n)$ is fixed. Here, while $\mathscr{D}$ contains $\mathscr{C}_0$, $\mathscr{Q}(\Sigma)$ is contained in the class $\mathscr{P}_{\max}(\Sigma)$. In other words, there is a trade-off between $\mathscr{D}$ and $\mathscr{Q}(\Sigma)$: the larger $\mathscr{D}$ is, the smaller $\mathscr{Q}(\Sigma)$ is. In fact, to evaluate the risk of $\hat{\beta}$ in $\mathscr{C}_1$, we need the existence of the second moments,

$$E_P\{\hat{\beta}'\hat{\beta}\} < \infty \qquad \text{for any} \quad \hat{\beta} \in \mathscr{D},$$

which is an additional restriction on the class $\mathscr{P}_{\max}(\Sigma)$.

LEMMA 5.1. *$\widetilde{\mathscr{C}}_{\mathrm{KT}} = \mathscr{C}_1$. In particular, if the distribution $P = \mathscr{L}(\varepsilon)$ satisfies $P(\varepsilon = 0) = 0$, then $\mathscr{C}_{\mathrm{KT}} = \mathscr{C}_1$ a.s. P, where $\mathscr{C}_{\mathrm{KT}}$ and $\widetilde{\mathscr{C}}_{\mathrm{KT}}$ are defined in* (2.6) *and* (2.5), *respectively.*

Proof. $\widetilde{\mathscr{C}}_{\mathrm{KT}} \subset \mathscr{C}_1$ follows because any $\tilde{\beta} = C(e)\, y + a\chi_{\{e=0\}} \in \widetilde{\mathscr{C}}_{\mathrm{KT}}$ can be expressed as

$$\begin{aligned}\tilde{\beta} &= C(e)\{X(X'X)^{-1}\, X' + N\}\; y + a\chi_{\{e=0\}}\\ &= (X'X)^{-1}\, X'y + \{C(e)\, e + a\chi_{\{e=0\}}\}\\ &\equiv b(I_n) + d(e) \qquad \text{(say)},\end{aligned} \tag{5.1}$$

where the matrix identity $X(X'X)^{-1}\, X' + N = I_n$ is used in the first line of (5.1). On the other hand, $\widetilde{\mathscr{C}}_{\mathrm{KT}} \supset \mathscr{C}_1$ follows because any $\hat{\beta} = b(I_n) + d(e) \in \mathscr{C}_1$ can be expressed as

$$\begin{aligned}\hat{\beta} &= b(I_n) + d(e)\{\chi_{\{e=0\}} + \chi_{\{e\neq 0\}}\}\\ &= (X'X)^{-1}\, X'y + \chi_{\{e\neq 0\}} d(e)(e'e)^{-1}\, e'e + d(e)\, \chi_{\{e=0\}}\\ &= \{(X'X)^{-1}\, X' + \chi_{\{e\neq 0\}} d(e)(e'e)^{-1}\, e'\}\; y + d(0)\, \chi_{\{e=0\}}\\ &\equiv C(e)\; y + a\chi_{\{e=0\}} \qquad \text{(say)},\end{aligned} \tag{5.2}$$

where the equality $e'y = e'e$ is used in the third line of (5.2). In particular, if P satisfies $P(\varepsilon = 0) = 0$, then $P(e = 0) = 0$ and hence $\mathscr{C}_{\mathrm{KT}} = \mathscr{C}_1$ a.s. by its definition. This completes the proof. ∎

It is noted that if P has a probability density function (pdf) with respect to Lebesgue measure in R^n, then $P(e = 0) = 0$ holds, implying $\mathscr{C}_1 = \mathscr{C}_{\mathrm{KT}}$ a.s. From the characterization of $\mathscr{C}_1$, it is easy to see that a GLSE $b(\hat{\Sigma})$ with $\hat{\Sigma} = \hat{\Sigma}(e)$ belongs to $\mathscr{C}_{\mathrm{KT}} \subset \mathscr{C}_1$ and that $\mathscr{C}_0 \subset \mathscr{C}_{\mathrm{KT}} \subset \mathscr{C}_1$.

Next we extend the result in Section 3 to specify $(\mathscr{D}, \mathscr{Q}(\Sigma))$ for which $b(\Sigma)$ is optimal. Let

$$\mathscr{S}_1(m, n) = \{\Phi(\cdot) \mid \Phi\colon R^m \to \mathscr{S}(n) \text{ is a measurable matrix-valued function}\}$$

and define an equivalence relation $\sim$ on $\mathscr{S}_1(n, n)$ by

$$\Phi(\cdot) \sim \Sigma(\cdot) \qquad \text{if and only if} \quad b(\Phi(e)) = b(\Sigma(e)) \qquad \text{a.s. } P.$$

Then the following results are straightforward from Theorem 3.1.

PROPOSITION 5.2. (1) *$\Phi(\cdot) \sim \Sigma(\cdot)$ if $\Phi(\cdot) \in \mathscr{R}_1(\Sigma(\cdot))$, where*

$$\mathscr{R}_1(\Sigma(\cdot)) = \{\Phi(\cdot) \in \mathscr{S}_1(n, n) \mid \Phi(e) = X\Upsilon(e) X' + \Sigma(e) Z\Delta(e) Z\Sigma(e)', \\ \Upsilon(\cdot) \in \mathscr{S}_1(n, k), \Delta(\cdot) \in \mathscr{S}_1(n, n-k)\}.$$

(2) *$\mathscr{C}_1 = \mathscr{B}_1$ with $\mathscr{B}_1 = \{b(\Phi(e)) \mid \Phi(\cdot) \in \mathscr{S}_1(n, n)\}$. In fact, as $\mathscr{C}_1 = \mathscr{C}_{\mathrm{KT}}$, $\hat{\beta}(y) = C(e)\, y$ is a GLSE with $\Phi(e)^{-1} = C(e)' C(e) + ZZ'$.*

Now $\hat{\beta} \in \mathscr{C}_1$ is expressed as

$$\begin{aligned} \hat{\beta} - \beta &= (X'X)^{-1} X'\{X(X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} + \Sigma Z(Z'\Sigma Z)^{-1} Z'\}\, \varepsilon + d(e) \\ &= (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}\varepsilon + \{(X'X)^{-1} X'\Sigma Z(Z'\Sigma Z)^{-1} Z'\varepsilon + d(e)\} \\ &= A^{-1/2}\tilde{\eta}_1 + h_{\hat{\beta}}(\tilde{\eta}_2) \qquad \text{(say)}, \end{aligned} \tag{5.3}$$

where the identity (3.7) is used in the first line of (5.3) and

$$\begin{aligned} h_{\hat{\beta}}(\tilde{\eta}_2) &= (X'X)^{-1} X'\Sigma Z(Z'\Sigma Z)^{-1} Z'\varepsilon + d(e) \\ &= (X'X)^{-1} X'\Sigma Z(Z'\Sigma Z)^{-1/2}\, \tilde{\eta}_2 + d(Z(Z'\Sigma Z)^{1/2}\, \tilde{\eta}_2). \end{aligned} \tag{5.4}$$

Here, it is noted that the OLS residual vector e in (2.4) is a function of $\tilde{\eta}_2$:

$$e = ZZ'\varepsilon = Z(Z'\Sigma Z)^{1/2}\, \tilde{\eta}_2.$$

For a GLSE $b(\hat{\Sigma})$ in (2.7), rewriting it as $b(\hat{\Sigma}) = C(e)\, y$ with $C(e) = (X'\hat{\Sigma}^{-1}X)^{-1} X'\hat{\Sigma}^{-1}$ yields $b(\hat{\Sigma}) = A^{-1/2}\tilde{\eta}_1 + h_{b(\hat{\Sigma})}(\tilde{\eta}_2)$ with

$$h_{b(\hat{\Sigma})}(\tilde{\eta}_2) = C(Z(Z'\Sigma Z)^{-1/2}\, \tilde{\eta}_2)\, \Sigma Z(Z\Sigma Z)^{-1/2}\, \tilde{\eta}_2.$$

The risk matrix of $\hat{\beta} \in \mathscr{C}_1$ is decomposed as

$$\begin{aligned} R_P(\hat{\beta}, \beta) = {} & A^{-1/2}E_P\{\tilde{\eta}_1\tilde{\eta}_1'\}\, A^{-1/2} + E_P\{h_{\hat{\beta}}(\tilde{\eta}_2)\, h_{\hat{\beta}}(\tilde{\eta}_2)'\} \\ & + A^{-1/2}E_P\{\tilde{\eta}_1 h_{\hat{\beta}}(\tilde{\eta}_2)'\} + E_P\{h_{\hat{\beta}}(\tilde{\eta}_2)\, \tilde{\eta}_1'\}\, A^{-1/2} \\ = {} & V_{11} + V_{22} + V_{12} + V_{21} \qquad \text{(say)}, \end{aligned} \tag{5.5}$$

as long as the four terms are finite. Here since $\mathscr{C}_1$ includes $\mathscr{C}_0$, it is necessary to assume $(\mu, \Phi) \in \mathscr{M}(\Sigma) \times \mathscr{R}(\Sigma)$ and to let $P = \mathscr{L}(\varepsilon)$ move over $\mathscr{P}_{\max}(\Sigma)$, so that the linear result in Section 4 should hold in our nonlinear extension.

Hence we impose the moment condition on $\mathscr{C}_1$: $E_P\{\hat{\beta}'\hat{\beta}\} < \infty$ for any $P \in \mathscr{P}_{\max}(\Sigma)$, which holds if and only if $\hat{\beta}$ belongs to

$$\mathscr{D}_1 = \{\hat{\beta} \in \mathscr{C}_1 \mid E_P\{\hat{\beta}'\hat{\beta}\} < \infty \text{ for any } P \in \mathscr{P}_{\max}(\Sigma)\}. \tag{5.6}$$

Clearly $\mathscr{D}_1$ is the maximal class of estimators in $\mathscr{C}_1$ which have the second moments for any $P \in \mathscr{P}_{\max}(\Sigma)$, and it contains all the GLSEs with second moments. For a sufficient condition for a GLSE to have finite second moments, see Kariya and Toyooka (1985).

THEOREM 5.3. *For a fixed $\Sigma \in \mathscr{S}(n)$, let*

$$\mathscr{Q}_1(\Sigma) = \{P \in \mathscr{P}_{\max}(\Sigma) \mid E_P\{\tilde{\eta}_1 h_{\hat{\beta}}(\tilde{\eta}_2)'\} = 0 \textit{ for any } \hat{\beta} \in \mathscr{D}_1\}. \tag{5.7}$$

Then

$$b(\Sigma) \textit{ is } (\mathscr{D}_1, \mathscr{Q}_1(\Sigma))\textit{-optimal.} \tag{5.8}$$

The proof is clear from (5.5) and the definitions of $\mathscr{D}_1$ and $\mathscr{Q}_1(\Sigma)$. Clearly $\mathscr{Q}_1(\Sigma) \subset \mathscr{P}_{\max}(\Sigma)$, although $\mathscr{D}_1 \supset \mathscr{C}_0$. Therefore the nonlinear version of the GMT in the above theorem is not completely stronger than the GMT in (2.2).

In other words, the risk matrix of he GME gives the lower bound for the risk matrices of estimators in $\mathscr{D}_1$ whenever $P \in \mathscr{Q}_1(\Sigma)$. Hence a measure of efficiency of an estimator $\hat{\beta}$ in $\mathscr{D}_1$ should be defined relative to the GME $b(\Sigma)$ in terms of the risk, when P belongs to $\mathscr{Q}_1(\Sigma)$. The measure depends on P in $\mathscr{Q}_1(\Sigma)$ as well as $\hat{\beta}$.

We describe two important subclasses of $\mathscr{Q}_1(\Sigma)$ in (5.7).

COROLLARY 5.4. *For a fixed $\Sigma \in \mathscr{S}(n)$, let*

$$\mathscr{Q}_2(\Sigma) = \{P \in \mathscr{P}_{\max}(\Sigma) \mid E_P\{\tilde{\eta}_1 \mid \tilde{\eta}_2\} = 0 \textit{ a.s. } \tilde{\eta}_2\}, \tag{5.9}$$

$$\mathscr{Q}_3(\Sigma) = \{P \in \mathscr{P}_{\max}(\Sigma) \mid \mathscr{L}_P(-\tilde{\eta}_1, \tilde{\eta}_2) = \mathscr{L}_P(\tilde{\eta}_1, \tilde{\eta}_2)\}, \tag{5.10}$$

where $\mathscr{L}_P(\cdot)$ denotes the distribution of $\cdot$ under P. Then

$$\mathscr{Q}_3(\Sigma) \subset \mathscr{Q}_2(\Sigma) \subset \mathscr{Q}_1(\Sigma) \tag{5.11}$$

holds, and hence

$$b(\Sigma) \textit{ is } (\mathscr{D}_1, \mathscr{Q}_i(\Sigma))\textit{-optimal} \qquad (i = 2, 3). \tag{5.12}$$

Proof. For any $P \in \mathscr{Q}_2(\Sigma)$ and any $\hat{\beta} \in \mathscr{D}_1$, it is easy to see that $E_P\{\tilde{\eta}_1 h(\tilde{\eta}_2)'\} = 0$ follows from $E_P\{\tilde{\eta}_1 \mid \tilde{\eta}_2\} = 0$, proving $\mathscr{Q}_2(\Sigma) \subset \mathscr{Q}_1(\Sigma)$. That $\mathscr{Q}_3(\Sigma) \subset \mathscr{Q}_2(\Sigma)$ is obvious. ∎

In the following we will show that $\mathscr{Q}_i(\Sigma)$'s $(i=2,3)$ include some interesting classes of distributions such as a class of elliptically symmetric distributions.

5.1. *Elliptically Symmetric Distributions*

In the model (1.1), suppose that $P=\mathscr{L}(\varepsilon)$ belongs to $\mathscr{E}(\mu,\Phi)$, the class of elliptically symmetric distributions with mean $\mu\in R^n$ and covariance matrix $\gamma\Phi$ for some $\gamma>0$. Here, $\mathscr{L}(\varepsilon)\in\mathscr{E}(\mu,\Phi)$ if and only if $\mathscr{L}(\varepsilon)\in\mathscr{P}(\mu,\Phi)$ and

$$\mathscr{L}(\Gamma\Phi^{-1/2}(\varepsilon-\mu))=\mathscr{L}(\Phi^{-1/2}(\varepsilon-\mu))$$

holds for any $n\times n$ orthogonal matrices Γ. For $\Sigma\in\mathscr{S}(n)$, a sufficient condition for $b(\Sigma)$ to be $(\mathscr{D}_1,\mathscr{E}(\mu,\Phi))$-optimal is that $\mathscr{E}(\mu,\Phi)\subset\mathscr{Q}_3(\Sigma)$.

THEOREM 5.5. *Assume $\mu\in L(\Sigma Z)$. Then $\mathscr{E}(\mu,\Phi)\subset\mathscr{Q}_3(\Sigma)$ holds if and only if $\Phi\in\mathscr{R}(\Sigma)$.*

Proof. Suppose first that $\mathscr{E}(\mu,\Phi)\subset\mathscr{Q}_3(\Sigma)$. Since $\mathscr{Q}_3(\Sigma)$ is a subclass of $\mathscr{P}_{\max}(\Sigma)$, the matrix Φ is clearly in $\mathscr{R}(\Sigma)$. Conversely, suppose $\Phi\in\mathscr{R}(\Sigma)$. Then $\mathscr{E}(\mu,\Phi)\subset\mathscr{P}_{\max}(\Sigma)$, and for any $P\in\mathscr{E}(\mu,\Phi)$, the distribution of $\tilde{\eta}=\Gamma_\Sigma\Sigma^{-1/2}\varepsilon=(\tilde{\eta}_1',\tilde{\eta}_2')'$ under P satisfies

$$\mathscr{L}_P(\tilde{\eta})=\mathscr{L}_P(\tilde{\eta}_1,\tilde{\eta}_2)\in\mathscr{E}(\tilde{\mu},Q),$$

where $\mu=\Sigma Zd$ with $d\in R^{n-k}$,

$$\begin{aligned}\tilde{\mu}&=\Gamma_\Sigma\Sigma^{-1/2}\mu\\&=\begin{pmatrix}(X'\Sigma^{-1}X)^{-1/2}\,X'\Sigma^{-1/2}\Sigma^{-1/2}\Sigma Zd\\(Z'\Sigma Z)^{-1/2}\,Z'\Sigma^{1/2}\Sigma^{-1/2}\Sigma Zd\end{pmatrix}\\&=\begin{pmatrix}0\\(Z'\Sigma Z)^{1/2}\,d\end{pmatrix}\end{aligned}\tag{5.13}$$

and

$$\begin{aligned}Q&=\Gamma_\Sigma\Sigma^{-1/2}\Phi\Sigma^{-1/2}\Gamma_\Sigma'\\&=\begin{pmatrix}(X'\Sigma^{-1}X)^{-1/2}\,X'\Sigma^{-1}\Phi\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1/2} & (X'\Sigma^{-1}X)^{-1/2}\,X'\Sigma^{-1}\Phi Z(Z'\Sigma Z)^{-1/2}\\(Z'\Sigma Z)^{-1/2}\,Z'\Phi\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1/2} & (Z'\Sigma Z)^{-1/2}\,Z'\Phi Z(Z'\Sigma Z)^{-1/2}\end{pmatrix}\\&=\begin{pmatrix}Q_{11}&Q_{12}\\Q_{21}&Q_{22}\end{pmatrix}\quad\text{(say).}\end{aligned}$$

Since $X'\Sigma^{-1}\Phi Z = 0$, this implies $Q_{12} = Q'_{21} = 0$, from which $\mathscr{L}_P(-\tilde{\eta}_1, \tilde{\eta}_2) = \mathscr{L}_P(\tilde{\eta}_1, \tilde{\eta}_2)$ follows, proving $P \in \mathscr{Q}_3(\Sigma)$. ▮

By this theorem, when $\mu \in L(\Sigma Z)$, an elliptically symmetric distribution is a member of $\mathscr{Q}_3(\Sigma)$ if and only if $\Phi \in \mathscr{R}(\Sigma)$.

COROLLARY 5.6. *For a fixed $\Sigma \in \mathscr{S}(n)$,*

$$b(\Sigma) \text{ is } (\mathscr{D}_1, \tilde{\mathscr{E}}(\Sigma))\text{-optimal},$$

where

$$\tilde{\mathscr{E}}(\Sigma) = \bigcup \{\mathscr{E}(\mu, \Phi) \mid (\mu, \Phi) \in L(\Sigma Z) \times \mathscr{R}(\Sigma)\}.$$

5.2. *Semi-elliptically Symmetric Distributions*

The second class is the set of distributions given by

$$\mathscr{E}_0(\mu, \Phi) = \{P \in \mathscr{P}(\mu, \Phi) \mid f_P \in \mathscr{F}(\mu, \Phi)\},$$

where $\mathscr{F}$ is the set of pdf's of the form

$$\begin{aligned} f_P(\varepsilon \mid \mu, \Phi) = |\Phi|^{-1/2}\, & q_1((\varepsilon - \mu)'\, \Phi^{-1/2} \bar{X}_\Phi \bar{X}'_\Phi \Phi^{-1/2}(\varepsilon - \mu)) \\ & \times q_2((\varepsilon - \mu)'\, \Phi^{-1/2} \bar{Z}_\Phi \bar{Z}'_\Phi \Phi^{-1/2}(\varepsilon - \mu)), \end{aligned}$$

for some nonnegative functions q_1 and q_2 on $[0, \infty)$ such that $\int_{R^k} q_1(x'_1 x_1)\, dx_1 = 1$ and $\int_{R^{n-k}} q_2(x'_2 x_2)\, dx_2 = 1$, where $\bar{X}_\Phi = \Phi^{-1/2} X (X'\Phi^{-1} X)^{-1/2} : n \times k$ and $\bar{Z}_\Phi = \Phi^{1/2} Z (Z'\Phi Z)^{-1/2} : n \times (n-k)$. Then the pdf of $\xi = \Gamma_\Phi \Phi^{-1/2} \varepsilon = (\xi'_1, \xi'_2)'$ with $\Gamma'_\Phi = (\bar{X}_\Phi, \bar{Z}_\Phi)$ is expressed as

$$\begin{aligned} f^*_P(\xi_1, \xi_2) = q_1((\xi_1 - \nu_1)'\, (\xi_1 - \nu_1))\, q_2((\xi_2 - \nu_2)'\, (\xi_2 - \nu_2)) \\ \text{with} \quad \nu = \Gamma_\Phi \Phi^{-1/2} \varepsilon = (\nu'_1, \nu'_2)'. \end{aligned}$$

That is, ξ_i's are independently and elliptically symmetrically distributed $(i = 1, 2)$.

Fix a matrix $\Sigma \in \mathscr{S}(n)$. For this family we still obtain the following result.

THEOREM 5.7. *Assume $\mu \in L(\Sigma Z)$. Then $\mathscr{E}_0(\mu, \Phi) \subset \mathscr{Q}_3(\Sigma)$ holds if and only if $\Phi \in \mathscr{R}(\Sigma)$.*

Proof. Suppose $\Phi \in \mathscr{R}(\Sigma)$. For any $P \in \mathscr{E}_0(\mu, \Phi)$, let $\tilde{f}_P(\tilde{\eta}_1, \tilde{\eta}_2)$ be the pdf of $\tilde{\eta} = \Gamma_\Sigma \Sigma^{-1/2}\varepsilon = (\tilde{\eta}_1', \tilde{\eta}_2')'$ under P. Then the function $\tilde{f}_P$ is calculated as

$$\tilde{f}_P(\tilde{\eta}_1, \tilde{\eta}_2) = |\Phi\Sigma^{-1}|^{-1/2}\, q_1((\tilde{\eta}-\tilde{\mu})'\, G(\tilde{\eta}-\tilde{\mu}))\, q_2((\tilde{\eta}-\tilde{\mu})'\, F(\tilde{\eta}-\tilde{\mu})),$$

where $\tilde{\mu} = \Gamma_\Sigma \Sigma^{-1/2}\mu = (0', \tilde{\mu}_2')'$ is the same as (5.13), and G and F are $n \times n$ symmetric matrices given by

$$\begin{aligned} G &= \Gamma_\Sigma \Sigma^{1/2}\Phi^{-1/2}\bar{X}_\Phi \bar{X}_\Phi' \Phi^{-1/2}\Sigma^{1/2}\Gamma_\Sigma' \\ &= \begin{pmatrix} (X'\Sigma^{-1}X)^{-1/2}\, X'\Phi^{-1}X(X'\Sigma^{-1}X)^{-1/2} & (X'\Sigma^{-1}X)^{-1/2}\, X'\Phi^{-1}\Sigma Z(Z'\Sigma Z)^{-1/2} \\ (Z'\Sigma Z)^{-1/2}\, Z'\Sigma\Phi^{-1}X(X'\Sigma^{-1}X)^{-1/2} & (Z'\Sigma Z)^{-1/2}\, Z'\Sigma\Phi^{-1}X(X'\Phi^{-1}X)^{-1}\, X'\Phi^{-1}\Sigma Z(Z'\Sigma Z)^{-1/2} \end{pmatrix} \\ &= \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \quad \text{(say)} \end{aligned}$$

and

$$\begin{aligned} H &= \Gamma_\Sigma \Sigma^{1/2}\Phi^{-1/2}\bar{Z}_\Phi \bar{Z}_\Phi' \Phi^{-1/2}\Sigma^{1/2}\Gamma_\Sigma' \\ &= \begin{pmatrix} 0 & 0 \\ 0 & (Z'\Sigma Z)^{1/2}\,(Z'\Phi Z)^{-1}\,(Z'\Sigma Z)^{1/2} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & H_{22} \end{pmatrix} \quad \text{(say)}, \end{aligned}$$

respectively. Since $X'\Phi^{-1}\Sigma Z = 0$ implies that G_{12}, G_{21}, and G_{22} are zero matrices, the pdf $\tilde{f}_P(\tilde{\eta}_1, \tilde{\eta}_2)$ is obtained as

$$\tilde{f}_P(\tilde{\eta}_1, \tilde{\eta}_2) = |\Phi\Sigma^{-1}|^{-1/2}\, q_1(\tilde{\eta}_1' G_{11}\tilde{\eta}_1)\, q_2((\tilde{\eta}_2-\tilde{\mu}_2)'\, H_{22}(\tilde{\eta}_2-\tilde{\mu}_2)),$$

from which $\mathscr{L}_P(-\tilde{\eta}_1, \tilde{\eta}_2) = \mathscr{L}_P(\tilde{\eta}_1, \tilde{\eta}_2)$ follows, proving $P \in \mathscr{Q}_3(\Sigma)$. Since the converse is clear, the proof is complete. ▮

COROLLARY 5.8. *For $\Sigma \in \mathscr{S}(n)$,*

$$b(\Sigma) \text{ is } (\mathscr{D}_1, \tilde{\mathscr{E}}_0(\Sigma))\text{-optimal}, \tag{5.14}$$

where

$$\tilde{\mathscr{E}}_0(\Sigma) = \bigcup \{\mathscr{E}_0(\mu, \Phi) \mid (\mu, \Phi) \in L(\Sigma Z) \times \mathscr{R}(\Sigma)\}. \tag{5.15}$$

5.3. *A Subclass of* $\mathcal{Q}_2(\Sigma)$

Finally we specify a subclass of $\mathcal{Q}_2(\Sigma)$ such that it is not necessarily contained in $\mathcal{Q}_3(\Sigma)$. Let

$$\mathcal{A}(\mu, \Sigma) = \{P \in \mathcal{P}(\mu, \Phi) \mid f_P \in \mathcal{F}(\mu, \Phi)\},$$

where f_P denotes the pdf of P, and $\mathcal{F}(\mu, \Phi)$ denotes the set of pdf's on R^n of the form

$$f_P(\varepsilon \mid \mu, \Phi) = |\Phi|^{-1/2}\, q(\Gamma_\Phi \Phi^{-1/2}(\varepsilon - \mu)),$$

where q is a nonnegative function on R^n such that $\int_{R^n} q(x)\, dx = 1$ and

$$\int_{R^k} x_1 q(x_1, x_2)\, dx_1 = 0 \qquad \text{a.e. } x_2 \qquad \text{with} \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

$$x_1 \colon k \times 1 \qquad \text{and} \qquad x_2 \colon (n-k) \times 1. \tag{5.16}$$

Let $\xi = \Gamma_\Phi \Phi^{-1/2}\varepsilon = (\xi_1', \xi_2')'$ and $\nu = \Gamma_\Phi \Phi^{-1/2}\mu = (\nu_1', \nu_2')'$. Then the pdf of $\mathcal{L}_P(\xi) = \mathcal{L}_P(\xi_1, \xi_2)$ is written as

$$f_P^*(\xi) = f_P^*(\xi_1, \xi_2) = q(\xi - \nu) = q(\xi_1 - \nu_1, \xi_2 - \nu_2),$$

which implies that $E_P\{\xi_1 \mid \xi_2\} = \nu_1$ a.s. ξ_2. Here, fix $\Sigma \in \mathcal{S}(n)$ and let $\tilde{\eta} = (\tilde{\eta}_1', \tilde{\eta}_2')' = \Gamma_\Sigma \Sigma^{-1/2}\varepsilon$. Since $\nu = 0$ and $\xi = \tilde{\eta}$ hold when $\mu = 0$ and $\Phi = \Sigma$, the zero integral restriction in (5.16) implies $E_P\{\tilde{\eta}_1 \mid \tilde{\eta}_2\} = E_P\{\xi_1 \mid \xi_2\} = 0$. Hence it is not restrictive.

THEOREM 5.9. *Assume* $\mu \in L(\Sigma Z)$. *Then* $\mathcal{A}(\mu, \Phi) \subset \mathcal{Q}_2(\Sigma)$ *holds if and only if* $\Phi \in \mathcal{R}(\Sigma)$.

Proof. Suppose that $\Phi \in \mathcal{R}(\Sigma)$, that is, Φ satisfies $X'\Phi^{-1}\Sigma Z = 0$. For any $P \in \mathcal{A}(\mu, \Phi)$, let $\tilde{f}_P(\tilde{\eta}_1, \tilde{\eta}_2)$ be the pdf of $\tilde{\eta} = (\tilde{\eta}_1', \tilde{\eta}_2')'$ under P. Then $\tilde{f}_P(\tilde{\eta}_1, \tilde{\eta}_2)$ is calculated as

$$\tilde{f}_P(\tilde{\eta}_1, \tilde{\eta}_2) = |\Phi\Sigma^{-1}|^{-1/2}\, q(G(\tilde{\eta} - \tilde{\mu})),$$

where $\tilde{\mu} = \Gamma_\Sigma \Sigma^{-1/2}\mu = (0', \tilde{\mu}_2')'$ is the same as (5.13), and G is an $n \times n$ nonsingular matrix given by

$$\begin{aligned}
Q &= \Gamma_\Phi \Phi^{-1/2} \Sigma^{1/2} \Gamma_\Sigma' \\
&= \begin{pmatrix} (X'\Phi^{-1}X)^{1/2}\,(X'\Sigma^{-1}X)^{-1/2} & (X'\Phi^{-1}X)^{1/2}\, X'\Phi^{-1}\Sigma Z (Z'\Sigma Z)^{-1/2} \\ 0 & (Z'\Phi Z)^{-1/2}\,(Z'\Sigma Z)^{1/2} \end{pmatrix} \\
&= \begin{pmatrix} Q_{11} & Q_{12} \\ 0 & Q_{22} \end{pmatrix} \qquad \text{(say)}.
\end{aligned}$$

Since $Q_{12}=0$ follows from $X'\Phi^{-1}\Sigma Z=0$, we have

$$\tilde{f}_P(\tilde{\eta}_1, \tilde{\eta}_2) = |\Phi\Sigma^{-1}|^{-1/2}\, q(G_{11}\tilde{\eta}_1, G_{22}(\tilde{\eta}_2 - \tilde{\mu}_2)),$$

from which $P \in \mathscr{Q}_2(\Sigma)$ follows. In fact,

$$\begin{aligned}\int_{R^k} \tilde{\eta}_1 \tilde{f}_P(\tilde{\eta}_1, \tilde{\eta}_2)\, d\tilde{\eta}_1 &= |\Phi\Sigma^{-1}|^{-1/2} \int_{R^k} \tilde{\eta}_1 q(G_{11}\tilde{\eta}_1, G_{22}(\tilde{\eta}_2 - \tilde{\mu}_2))\, d\tilde{\eta}_1 \\ &= |\Phi\Sigma^{-1}|^{-1/2}\, |G_{11}|^{-1}\, G_{11}^{-1} \int_{R^k} x_1 q(x_1, x_2)\, dx_1 \\ &= 0 \qquad \text{a.e. } x_2.\end{aligned}$$

implies $E_P\{\tilde{\eta}_1 \mid \tilde{\eta}_2\} = 0$ a.s. $\tilde{\eta}_2$, where $x_1 = G_{11}\tilde{\eta}_1$ and $x_2 = G_{22}(\tilde{\eta}_2 - \tilde{\mu}_2)$. The converse is clear. ∎

COROLLARY 5.10. *For $\Sigma \in \mathscr{S}(n)$,*

$$b(\Sigma) \text{ is } (\mathscr{D}_1, \tilde{\mathscr{A}}(\Sigma))\text{-optimal},$$

where

$$\tilde{\mathscr{A}}(\Sigma) = \bigcup \{\mathscr{A}(\mu, \Phi) \mid (\mu, \Phi) \in L(\Sigma Z) \times \mathscr{R}(\Sigma)\}.$$

ACKNOWLEDGMENT

The authors express their sincere gratitude to a referee for his valuable comments.

REFERENCES

1. M. M. Ali and R. Ponnapalli, An optimality of the Gauss–Markov estimator, *J. Multivariate Anal.* **32** (1990), 171–176.
2. D. W. K. Andrews and P. C. B. Phillips, Best median-unbiased estimation in linear regression with bounded asymmetric loss functions, *J. Amer. Statist. Assoc.* **82** (1987), 886–893.
3. R. Berk and J. T. Hwang, Optimality of the least squares estimator, *J. Multivariate Anal.* **30** (1989), 245–254.
4. M. L. Eaton, The Gauss-Markov Theorem in multivariate analysis, *in* "Multivariate Analysis" (P. R. Krishnaiah, Ed.), Vol. 6, pp. 177–201, North-Holland, Amsterdam, 1985.
5. M. L. Eaton, A characterization of spherical distributions, *J. Multivariate Anal.* **20** (1986), 272–276.
6. M. L. Eaton, Concentration inequarities for Gauss–Markov estimators, *J. Multivariate Anal.* **25** (1988), 119–138.
7. J. T. Hwang, Universal domination and stochastic domination: Estimation simultaneously under a broad class of loss functions, *Ann. Statist.* **13** (1985), 295–314.

8. D. R. Jensen, Symmetry and unimodality in linear inference, *J. Multivariate Anal.* **60** (1996), 188–202.
9. T. Kariya, A nonlinear version of the Gauss–Markov theorem, *J. Amer. Statist. Assoc.* **76** (1985), 476–477.
10. T. Kariya and Y. Toyooka, Nonlinear versions of the Gauss–Markov theorem and GLSE, *in* "Multivariate Analysis" (P. R. Krishnaiah, Ed.), Vol. 6, pp. 345–354, North-Holland, Amsterdam, 1985.
11. Y. G. Kuritsyn, On the least-squares method for elliptically contoured distributions, *Theory Prob. Appl.* **31** (1986), 738–740.
12. C. R. Rao and S. K. Mitra, "Generalized Inverse of Matrices and Its Applications," Wiley, New York, 1971.