BLUE: Robustness of LSE against normality

We now study properties of $l^\tau\widehat{\beta}$ and $\widehat{\sigma}^2$ under assumption A2, i.e., without the normality assumption on ε.

- From Theorem 3.6 and the proof of Theorem 3.7(ii), $l^\tau\widehat{\beta}$ (with an $l \in \mathscr{R}(Z)$) and $\widehat{\sigma}^2$ are still unbiased without normality.
- In what sense are $l^\tau\widehat{\beta}$ and $\widehat{\sigma}^2$ optimal beyond being unbiased?
- Some discussion about $\widehat{\sigma}^2$ can be found in Rao (1973, p. 228).

In general,

$$\mathrm{Var}(l^\tau\widehat{\beta}) = l^\tau(Z^\tau Z)^- Z^\tau \mathrm{Var}(\varepsilon) Z(Z^\tau Z)^- l.$$

If $l \in \mathscr{R}(Z)$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$ (assumption A2), then by the property of generalized inverse matrix,

$$\mathrm{Var}(l^\tau\widehat{\beta}) = \sigma^2 l^\tau(Z^\tau Z)^- l$$

which attains the Cramér-Rao lower bound under assumption A1.

We have the following result for the LSE $l^\tau\widehat{\beta}$.

Theorem 3.9

Consider linear model

$$X = Z\beta + \varepsilon. \tag{1}$$

with assumption A2.

(i) A necessary and sufficient condition for the existence of a linear unbiased estimator of $l^\tau\beta$ (i.e., an unbiased estimator that is linear in X) is $l \in \mathscr{R}(Z)$.

(ii) (Gauss-Markov theorem). If $l \in \mathscr{R}(Z)$, then the LSE $l^\tau\widehat{\beta}$ is the *best linear unbiased estimator* (BLUE) of $l^\tau\beta$ in the sense that it has the minimum variance in the class of linear unbiased estimators of $l^\tau\beta$.

Proof of (i)

The sufficiency has been established in Theorem 3.6.
Suppose now a linear function of X, $c^\tau X$ with $c \in \mathscr{R}^n$, is unbiased for $l^\tau\beta$. Then

$$l^\tau\beta = E(c^\tau X) = c^\tau EX = c^\tau Z\beta.$$

Since this equality holds for all β, $l = Z^\tau c$, i.e., $l \in \mathscr{R}(Z)$.

Proof of (ii)

Let $l \in \mathscr{R}(Z) = \mathscr{R}(Z^\tau Z)$.

Then $l = (Z^\tau Z)\zeta$ for some ζ and $l^\tau \widehat{\beta} = \zeta^\tau (Z^\tau Z)\widehat{\beta} = \zeta^\tau Z^\tau X$.

Let $c^\tau X$ be any linear unbiased estimator of $l^\tau \beta$.

From the proof of (i), $Z^\tau c = l$, and

$$\begin{aligned}
\mathrm{Cov}(\zeta^\tau Z^\tau X, c^\tau X - \zeta^\tau Z^\tau X) &= E(X^\tau Z\zeta c^\tau X) - E(X^\tau Z\zeta\zeta^\tau Z^\tau X) \\
&= \sigma^2 \mathrm{tr}(Z\zeta c^\tau) + \beta^\tau Z^\tau Z\zeta c^\tau Z\beta \\
&- \ \sigma^2 \mathrm{tr}(Z\zeta\zeta^\tau Z^\tau) - \beta^\tau Z^\tau Z\zeta\zeta^\tau Z^\tau Z\beta \\
&= \sigma^2 \zeta^\tau l + (l^\tau \beta)^2 - \sigma^2 \zeta^\tau l - (l^\tau \beta)^2 \\
&= 0.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathrm{Var}(c^\tau X) = & \ \mathrm{Var}(c^\tau X - \zeta^\tau Z^\tau X + \zeta^\tau Z^\tau X) \\
= & \ \mathrm{Var}(c^\tau X - \zeta^\tau Z^\tau X) + \mathrm{Var}(\zeta^\tau Z^\tau X) \\
+ & \ 2\,\mathrm{Cov}(\zeta^\tau Z^\tau X, c^\tau X - \zeta^\tau Z^\tau X) \\
= & \ \mathrm{Var}(c^\tau X - \zeta^\tau Z^\tau X) + \mathrm{Var}(l^\tau \widehat{\beta}) \\
\geq & \ \mathrm{Var}(l^\tau \widehat{\beta}).
\end{aligned}$$

Robustness of LSE against violation of $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$

Consider linear model (5) under assumption A3 ($E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon)$ is an unknown matrix).

An interesting question is under what conditions on $\mathrm{Var}(\varepsilon)$ is the LSE of $l^\tau \beta$ with $l \in \mathscr{R}(Z)$ still the BLUE.

If $l^\tau \widehat{\beta}$ is still the BLUE, then we say that $l^\tau \widehat{\beta}$, considered as a BLUE, is *robust* against violation of assumption A2.

A procedure having certain properties under an assumption is said to be robust against violation of the assumption iff the procedure still has the same properties when the assumption is (slightly) violated.

For example, the LSE of $l^\tau \beta$ with $l \in \mathscr{R}(Z)$, as an unbiased estimator, is robust against violation of assumption A1 or A2, since the LSE is unbiased as long as $E(\varepsilon) = 0$, which can be always assumed.

On the other hand, the LSE as a UMVUE may not be robust against violation of assumption A1.

Theorem 3.10

Consider model (5) with assumption A3.
The following are equivalent.

(a) $l^\tau \widehat{\beta}$ is the BLUE of $l^\tau \beta$ for any $l \in \mathscr{R}(Z)$.

(b) $E(l^\tau \widehat{\beta} \eta^\tau X) = 0$ for any $l \in \mathscr{R}(Z)$ and any η such that $E(\eta^\tau X) = 0$.

(c) $Z^\tau \operatorname{Var}(\varepsilon) U = 0$, where U is a matrix such that $Z^\tau U = 0$ and $\mathscr{R}(U^\tau) + \mathscr{R}(Z^\tau) = \mathscr{R}^n$.

(d) $\operatorname{Var}(\varepsilon) = Z \Lambda_1 Z^\tau + U \Lambda_2 U^\tau$ for some Λ_1 and Λ_2.

(e) The matrix $Z(Z^\tau Z)^- Z^\tau \operatorname{Var}(\varepsilon)$ is symmetric.

Proof

We first show that (a) and (b) are equivalent, which is an analogue of Theorem 3.2(i).
Suppose that (b) holds.
Let $l \in \mathscr{R}(Z)$.
If $c^\tau X$ is unbiased for $l^\tau \beta$, then $E(\eta^\tau X) = 0$ with $\eta = c - Z(Z^\tau Z)^- l$.
Hence, (b) implies (a) because

Proof (continued)

$$\begin{aligned}
\mathrm{Var}(c^\tau X) &= \mathrm{Var}(c^\tau X - l^\tau \widehat{\beta} + l^\tau \widehat{\beta}) \\
&= \mathrm{Var}(c^\tau X - l^\tau (Z^\tau Z)^- Z^\tau X + l^\tau \widehat{\beta}) \\
&= \mathrm{Var}(\eta^\tau X + l^\tau \widehat{\beta}) \\
&= \mathrm{Var}(\eta^\tau X) + \mathrm{Var}(l^\tau \widehat{\beta}) + 2\,\mathrm{Cov}(\eta^\tau X, l^\tau \widehat{\beta}) \\
&= \mathrm{Var}(\eta^\tau X) + \mathrm{Var}(l^\tau \widehat{\beta}) + 2E(l^\tau \widehat{\beta} \eta^\tau X) \\
&= \mathrm{Var}(\eta^\tau X) + \mathrm{Var}(l^\tau \widehat{\beta}) \\
&\geq \mathrm{Var}(l^\tau \widehat{\beta}).
\end{aligned}$$

Suppose now that there are $l \in \mathscr{R}(Z)$ and η such that $E(\eta^\tau X) = 0$ but $\delta = E(l^\tau \widehat{\beta} \eta^\tau X) \neq 0$.
Let $c_t = t\eta + Z(Z^\tau Z)^- l$.
From the previous proof,

$$\mathrm{Var}(c_t^\tau X) = t^2\,\mathrm{Var}(\eta^\tau X) + \mathrm{Var}(l^\tau \widehat{\beta}) + 2\delta t.$$

As long as $\delta \neq 0$, there exists a t such that $\mathrm{Var}(c_t^\tau X) < \mathrm{Var}(l^\tau \widehat{\beta})$.

Proof (continued)

This shows that $l^\tau \widehat{\beta}$ cannot be a BLUE and, therefore, (a) implies (b).

Next, we show that (b) implies (c).
Suppose that (b) holds.
Since $l \in \mathscr{R}(Z)$, $l = Z^\tau \gamma$ for some γ.
Let $\eta \in \mathscr{R}(U^\tau)$.
Then $E(\eta^\tau X) = \eta^\tau Z\beta = 0$ and, hence,

$$0 = E(l^\tau \widehat{\beta}\eta^\tau X) = E[\gamma^\tau Z(Z^\tau Z)^- Z^\tau XX^\tau \eta] = \gamma^\tau Z(Z^\tau Z)^- Z^\tau \,\mathrm{Var}(\varepsilon)\eta.$$

Since this equality holds for all $l \in \mathscr{R}(Z)$, it holds for all γ.
Thus,

$$Z(Z^\tau Z)^- Z^\tau \,\mathrm{Var}(\varepsilon)U = 0,$$

which implies

$$Z^\tau Z(Z^\tau Z)^- Z^\tau \,\mathrm{Var}(\varepsilon)U = Z^\tau \,\mathrm{Var}(\varepsilon)U = 0,$$

since $Z^\tau Z(Z^\tau Z)^- Z^\tau = Z^\tau$.
Thus, (c) holds.

Proof (continued)

Next, we show that (c) implies (d).
We need to use the following facts from the theory of linear algebra: there exists a nonsingular matrix C such that $\mathrm{Var}(\varepsilon) = CC^\tau$ and $C = ZC_1 + UC_2$ for some matrices C_j (since $\mathscr{R}(U^\tau) + \mathscr{R}(Z^\tau) = \mathscr{R}^n$).
Let $\Lambda_1 = C_1 C_1^\tau$, $\Lambda_2 = C_2 C_2^\tau$, and $\Lambda_3 = C_1 C_2^\tau$.
Then

$$\mathrm{Var}(\varepsilon) = Z\Lambda_1 Z^\tau + U\Lambda_2 U^\tau + Z\Lambda_3 U^\tau + U\Lambda_3^\tau Z^\tau \tag{2}$$

and $Z^\tau \, \mathrm{Var}(\varepsilon) U = Z^\tau Z \Lambda_3 U^\tau U$, which is 0 if (c) holds.
Hence, (c) implies

$$0 = Z(Z^\tau Z)^- Z^\tau Z \Lambda_3 U^\tau U (U^\tau U)^- U^\tau = Z\Lambda_3 U^\tau,$$

which with (2) implies (d).

We now show that (d) implies (e).
If (d) holds, then $Z(Z^\tau Z)^- Z^\tau \, \mathrm{Var}(\varepsilon) = Z\Lambda_1 Z^\tau$, which is symmetric.
Hence (d) implies (e).

To complete the proof, we need to show that (e) implies (b), which is left as an exercise.

As a corollary of this theorem, the following result shows when the UMVUE's in model (5) with assumption A1 are robust against the violation of $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$.

Corollary 3.3

Consider model (5) with a full rank Z, $\varepsilon = N_n(0, \Sigma)$, and an unknown positive definite matrix Σ.
Then $l^\tau \widehat{\beta}$ is a UMVUE of $l^\tau \beta$ for any $l \in \mathscr{R}^p$ iff one of (b)-(e) in Theorem 3.10 holds.

Example 3.16

Consider model (5) with β replaced by a random vector $\vec{\beta}$ that is independent of ε.
Such a model is called a linear model with random coefficients.
Suppose that $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$ and $E(\vec{\beta}) = \beta$.
Then

$$X = Z\beta + Z(\vec{\beta} - \beta) + \varepsilon = Z\beta + e, \tag{3}$$

where $e = Z(\vec{\beta} - \beta) + \varepsilon$ satisfies $E(e) = 0$ and

$$\mathrm{Var}(e) = Z\,\mathrm{Var}(\vec{\beta})Z^\tau + \sigma^2 I_n.$$

Since

$$Z(Z^\tau Z)^- Z^\tau\,\mathrm{Var}(e) = Z\,\mathrm{Var}(\vec{\beta})Z^\tau + \sigma^2 Z(Z^\tau Z)^- Z^\tau$$

is symmetric, by Theorem 3.10, the LSE $l^\tau\widehat{\beta}$ under model (3) is the BLUE for any $l^\tau\beta$, $l \in \mathscr{R}(Z)$.

If Z is of full rank and ε is normal, then, by Corollary 3.3, $l^\tau\widehat{\beta}$ is the UMVUE of $l^\tau\beta$ for any $l \in \mathscr{R}^p$.

Example 3.17 (Random effects models)

Suppose that

$$X_{ij} = \mu + A_i + e_{ij}, \quad j = 1,...,n_i, i = 1,...,m, \tag{4}$$

where $\mu \in \mathscr{R}$ is an unknown parameter, A_i's are i.i.d. random variables having mean 0 and variance σ_a^2, e_{ij}'s are i.i.d. random errors with mean 0 and variance σ^2, and A_i's and e_{ij}'s are independent.

Model (4) is called a one-way *random effects* model and A_i's are unobserved random effects.

Model (4) is a special case of model (5) with $\varepsilon_{ij} = A_i + e_{ij}$ and

$$\mathrm{Var}(\varepsilon) = \sigma_a^2 \Sigma + \sigma^2 I_n,$$

where Σ is a block diagonal matrix whose ith block is $J_{n_i} J_{n_i}^\tau$ and J_k is the k-vector of ones.

Under this model, $Z = J_n$, $n = \sum_{i=1}^m n_i$, and $Z(Z^\tau Z)^- Z^\tau = n^{-1} J_n J_n^\tau$.

Note that

$$J_n J_n^\tau \Sigma = \begin{pmatrix} n_1 J_{n_1} J_{n_1}^\tau & n_2 J_{n_1} J_{n_2}^\tau & \cdots & n_m J_{n_1} J_{n_m}^\tau \\ n_1 J_{n_2} J_{n_1}^\tau & n_2 J_{n_2} J_{n_2}^\tau & \cdots & n_m J_{n_2} J_{n_m}^\tau \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ n_1 J_{n_m} J_{n_1}^\tau & n_2 J_{n_m} J_{n_2}^\tau & \cdots & n_m J_{n_m} J_{n_m}^\tau \end{pmatrix},$$

which is symmetric if and only if $n_1 = n_2 = \cdots = n_m$.

Since $J_n J_n^\tau \mathrm{Var}(\varepsilon)$ is symmetric if and only if $J_n J_n^\tau \Sigma$ is symmetric, a necessary and sufficient condition for the LSE of μ to be the BLUE is that all n_i's are the same.

This condition is also necessary and sufficient for the LSE of μ to be the UMVUE when ε_{ij}'s are normal.

In some cases, we are interested in some (not all) linear functions of β. For example, consider $l^\tau \beta$ with $l \in \mathscr{R}(H)$, where H is an $n \times p$ matrix such that $\mathscr{R}(H) \subset \mathscr{R}(Z)$.

Proposition 3.4

Consider model (5) with assumption A3.
Suppose that H is a matrix such that $\mathscr{R}(H) \subset \mathscr{R}(Z)$.
A necessary and sufficient condition for the LSE $l^\tau \widehat{\beta}$ to be the BLUE of $l^\tau \beta$ for any $l \in \mathscr{R}(H)$ is $H(Z^\tau Z)^- Z^\tau \operatorname{Var}(\varepsilon) U = 0$, where U is the same as that in (c) of Theorem 3.10.

Example 3.18

Consider model (5) with assumption A3 and $Z = (H_1 \;\; H_2)$, where $H_1^\tau H_2 = 0$.
Suppose that under the reduced model

$$X = H_1 \beta_1 + \varepsilon,$$

$l^\tau \widehat{\beta}_1$ is the BLUE for any $l^\tau \beta_1$, $l \in \mathscr{R}(H_1)$

Example 3.18 (continued)

and that under the reduced model

$$X = H_2\beta_2 + \varepsilon,$$

$l^\tau\widehat{\beta}_2$ is not a BLUE for some $l^\tau\beta_2$, $l \in \mathscr{R}(H_2)$, where $\beta = (\beta_1, \beta_2)$ and $\widehat{\beta}_j$'s are LSE's under the reduced models.
Let $H = (H_1 \; 0)$ be $n \times p$.
Note that

$$H(Z^\tau Z)^- Z^\tau \operatorname{Var}(\varepsilon) U = H_1(H_1^\tau H_1)^- H_1^\tau \operatorname{Var}(\varepsilon) U,$$

which is 0 by Theorem 3.10 for the U given in (c) of Theorem 3.10, and

$$Z(Z^\tau Z)^- Z^\tau \operatorname{Var}(\varepsilon) U = H_2(H_2^\tau H_2)^- H_2^\tau \operatorname{Var}(\varepsilon) U,$$

which is not 0 by Theorem 3.10.
This implies that some LSE $l^\tau\widehat{\beta}$ is not a BLUE of $l^\tau\beta$ but $l^\tau\widehat{\beta}$ is the BLUE of $l^\tau\beta$ if $l \in \mathscr{R}(H)$.

Asymptotic properties of LSE

Without normality on ε or $\text{Var}(\varepsilon) = \sigma^2 I_n$, properties of the LSE may be derived under the asymptotic framework.

Theorem 3.11 (Consistency)

Consider model

$$X = Z\beta + \varepsilon \tag{5}$$

under assumption A3 ($E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon)$ is an unknown matrix).
Consider the LSE $l^\tau \widehat{\beta}$ with $l \in \mathscr{R}(Z)$ for every n.
Suppose that $\sup_n \lambda_+[\mathrm{Var}(\varepsilon)] < \infty$, where $\lambda_+[A]$ is the largest eigenvalue of the matrix A, and that $\lim_{n\to\infty} \lambda_+[(Z^\tau Z)^-] = 0$.
Then $l^\tau \widehat{\beta}$ is consistent in mse for any $l \in \mathscr{R}(Z)$.

Proof

The result follows from the fact that $l^\tau \widehat{\beta}$ is unbiased and

$$\mathrm{Var}(l^\tau \widehat{\beta}) = l^\tau (Z^\tau Z)^- Z^\tau \,\mathrm{Var}(\varepsilon) Z (Z^\tau Z)^- l \le \lambda_+[\mathrm{Var}(\varepsilon)] l^\tau (Z^\tau Z)^- l.$$

Without the normality assumption on ε, the exact distribution of $l^\tau\widehat{\beta}$ is very hard to obtain.
The asymptotic distribution of $l^\tau\widehat{\beta}$ is derived in the following result.

Theorem 3.12

Consider model (5) with assumption A3.
Suppose that $0 < \inf_n \lambda_-[\,\mathrm{Var}(\varepsilon)]$, where $\lambda_-[A]$ is the smallest eigenvalue of the matrix A, and that

$$\lim_{n\to\infty} \max_{1\le i\le n} Z_i^\tau (Z^\tau Z)^- Z_i = 0. \tag{6}$$

Suppose further that $n = \sum_{j=1}^k m_j$ for some integers k, m_j, $j = 1,...,k$, with m_j's bounded by a fixed integer m, $\varepsilon = (\xi_1,...,\xi_k)$, $\xi_j \in \mathscr{R}^{m_j}$, and ξ_j's are independent.

(i) If $\sup_i E|\varepsilon_i|^{2+\delta} < \infty$, then for any $l \in \mathscr{R}(Z)$,

$$l^\tau(\widehat{\beta}-\beta)\Big/\sqrt{\mathrm{Var}(l^\tau\widehat{\beta})} \to_d N(0,1). \tag{7}$$

(ii) Result (7) holds for any $l \in \mathscr{R}(Z)$ if, when $m_i = m_j$, $1 \le i < j \le k$, ξ_i and ξ_j have the same distribution.

Proof

For $l \in \mathscr{R}(Z)$,

$$l^\tau (Z^\tau Z)^- Z^\tau Z\beta - l^\tau \beta = 0$$

and

$$l^\tau(\widehat{\beta} - \beta) = l^\tau (Z^\tau Z)^- Z^\tau \varepsilon = \sum_{j=1}^{k} c_{nj}^\tau \xi_j,$$

where c_{nj} is the m_j-vector whose components are $l^\tau (Z^\tau Z)^- Z_i$, $i = k_{j-1}+1, ..., k_j$, $k_0 = 0$, and $k_j = \sum_{t=1}^{j} m_t$, $j = 1, ..., k$.
Note that

$$\sum_{j=1}^{k} \|c_{nj}\|^2 = l^\tau (Z^\tau Z)^- Z^\tau Z (Z^\tau Z)^- l = l^\tau (Z^\tau Z)^- l. \qquad (8)$$

Also,

$$\begin{aligned} \max_{1\le j\le k} \|c_{nj}\|^2 &\le m \max_{1\le i\le n} [l^\tau (Z^\tau Z)^- Z_i]^2 \\ &\le m l^\tau (Z^\tau Z)^- l \max_{1\le i\le n} Z_i^\tau (Z^\tau Z)^- Z_i, \end{aligned}$$

which, together with (8) and condition (6), implies that

Proof (continued)

$$\lim_{n\to\infty}\left(\max_{1\le j\le k}\|c_{nj}\|^2 \Big/ \sum_{j=1}^{k}\|c_{nj}\|^2\right) = 0.$$

The results then follow from Corollary 1.3.

Remarks

- Under the conditions of Theorem 3.12, $\mathrm{Var}(\varepsilon)$ is a diagonal block matrix with $\mathrm{Var}(\xi_j)$ as the jth diagonal block, which includes the case of independent ε_i's as a special case.
- Exercise 80 shows that condition (6) is almost a necessary condition for the consistency of the LSE.

Lemma 3.3

The following are sufficient conditions for (6).

(a) $\lambda_+[(Z^\tau Z)^-] \to 0$ and $Z_n^\tau (Z^\tau Z)^- Z_n \to 0$, as $n \to \infty$.

(b) There is an increasing sequence $\{a_n\}$ such that $a_n \to \infty$, $a_n/a_{n+1} \to 1$, and $Z^\tau Z/a_n$ converges to a positive definite matrix.