

Lecture 37: Robustness of LSE's

Consider model

$$X = Z\beta + \varepsilon. \quad (1)$$

under assumption A3 ($E(\varepsilon) = 0$ and $\text{Var}(\varepsilon)$ is an unknown matrix).

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

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

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

For example, the LSE of $l^T\beta$ with $l \in \mathcal{R}(Z)$, 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 without loss of generality.

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

Theorem 3.10. Consider model (1) with assumption A3. The following are equivalent.

- (a) $l^T\hat{\beta}$ is the BLUE of $l^T\beta$ for any $l \in \mathcal{R}(Z)$.
- (b) $E(l^T\hat{\beta}\eta^T X) = 0$ for any $l \in \mathcal{R}(Z)$ and any η such that $E(\eta^T X) = 0$.
- (c) $Z^T \text{Var}(\varepsilon)U = 0$, where U is a matrix such that $Z^T U = 0$ and $\mathcal{R}(U^T) + \mathcal{R}(Z^T) = \mathcal{R}^n$.
- (d) $\text{Var}(\varepsilon) = Z\Lambda_1 Z^T + U\Lambda_2 U^T$ for some Λ_1 and Λ_2 .
- (e) The matrix $Z(Z^T Z)^- Z^T \text{Var}(\varepsilon)$ 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 $l \in \mathcal{R}(Z)$.

If $c^T X$ is unbiased for $l^T\beta$, then $E(\eta^T X) = 0$ with $\eta = c - Z(Z^T Z)^- l$.

Hence

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

Suppose now that there are $l \in \mathcal{R}(Z)$ and η such that $E(\eta^T X) = 0$ but $\delta = E(l^T\hat{\beta}\eta^T X) \neq 0$.

Let $c_t = t\eta + Z(Z^T Z)^- l$.

From the previous proof,

$$\text{Var}(c_t^T X) = t^2 \text{Var}(\eta^T X) + \text{Var}(l^T\hat{\beta}) + 2\delta t.$$

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

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

We next show that (b) implies (c).

Suppose that (b) holds.

Since $l \in \mathcal{R}(Z)$, $l = Z^\tau \gamma$ for some γ .

Let $\eta \in \mathcal{R}(U^\tau)$.

Then $E(\eta^\tau X) = \eta^\tau Z\beta = 0$ and, hence,

$$0 = E(l^\tau \hat{\beta} \eta^\tau X) = E[\gamma^\tau Z(Z^\tau Z)^{-1} Z^\tau X X^\tau \eta] = \gamma^\tau Z(Z^\tau Z)^{-1} Z^\tau \text{Var}(\varepsilon) \eta.$$

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

Thus,

$$Z(Z^\tau Z)^{-1} Z^\tau \text{Var}(\varepsilon) U = 0,$$

which implies

$$Z^\tau Z(Z^\tau Z)^{-1} Z^\tau \text{Var}(\varepsilon) U = Z^\tau \text{Var}(\varepsilon) U = 0,$$

since $Z^\tau Z(Z^\tau Z)^{-1} Z^\tau = Z^\tau$.

Thus, (c) holds.

To 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 $\text{Var}(\varepsilon) = CC^\tau$ and $C = ZC_1 + UC_2$ for some matrices C_j (since $\mathcal{R}(U^\tau) + \mathcal{R}(Z^\tau) = \mathcal{R}^n$).

Let $\Lambda_1 = C_1 C_1^\tau$, $\Lambda_2 = C_2 C_2^\tau$, and $\Lambda_3 = C_1 C_2^\tau$.

Then

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

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

Hence, (c) implies

$$0 = Z(Z^\tau Z)^{-1} Z^\tau Z\Lambda_3 U^\tau U (U^\tau U)^{-1} U^\tau = Z\Lambda_3 U^\tau,$$

which with (2) implies (d).

If (d) holds, then $Z(Z^\tau Z)^{-1} Z^\tau \text{Var}(\varepsilon) = Z\Lambda_1 Z^\tau$, 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 (1) with assumption A1 are robust against the violation of $\text{Var}(\varepsilon) = \sigma^2 I_n$.

Corollary 3.3. Consider model (1) with a full rank Z , $\varepsilon = N_n(0, \Sigma)$, and an unknown positive definite matrix Σ . Then $l^\tau \hat{\beta}$ is a UMVUE of $l^\tau \beta$ for any $l \in \mathcal{R}^p$ if and only if one of (b)-(e) in Theorem 3.10 holds.

Example 3.16. Consider model (1) with β replaced by a random vector β that is independent of ε .

Such a model is called a linear model with random coefficients. Suppose that $\text{Var}(\varepsilon) = \sigma^2 I_n$ and $E(\boldsymbol{\beta}) = \beta$. Then

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

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

$$\text{Var}(e) = Z\text{Var}(\boldsymbol{\beta})Z^\tau + \sigma^2 I_n.$$

Since

$$Z(Z^\tau Z)^- Z^\tau \text{Var}(e) = Z\text{Var}(\boldsymbol{\beta})Z^\tau + \sigma^2 Z(Z^\tau Z)^- Z^\tau$$

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

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

Example 3.17 (Random effects models). Suppose that

$$X_{ij} = \mu + A_i + e_{ij}, \quad j = 1, \dots, n_i, i = 1, \dots, m, \quad (4)$$

where $\mu \in \mathcal{R}$ 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.

Let $\varepsilon_{ij} = A_i + e_{ij}$.

Then (4) is a special case of the general model (1) with

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

where Σ is a block diagonal matrix whose i th 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

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

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

Since $J_n J_n^\tau \text{Var}(\varepsilon)$ 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 $l \in \mathcal{R}(H)$, where H is an $n \times p$ matrix such that $\mathcal{R}(H) \subset \mathcal{R}(Z)$.

Proposition 3.4. Consider model (1) with assumption A3. Suppose that H is a matrix such that $\mathcal{R}(H) \subset \mathcal{R}(Z)$. A necessary and sufficient condition for the LSE $l^\tau \hat{\beta}$ to be the BLUE of $l^\tau \beta$ for any $l \in \mathcal{R}(H)$ is $H(Z^\tau Z)^- Z^\tau \text{Var}(\varepsilon)U = 0$, where U is the same as that in (c) of Theorem 3.10.

Example 3.18. Consider model (1) with assumption A3 and $Z = (H_1 \ H_2)$, where $H_1^\tau H_2 = 0$.

Suppose that under the reduced model

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

$l^\tau \hat{\beta}_1$ is the BLUE for any $l^\tau \beta_1$, $l \in \mathcal{R}(H_1)$, and that under the reduced model

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

$l^\tau \hat{\beta}_2$ is not a BLUE for some $l^\tau \beta_2$, $l \in \mathcal{R}(H_2)$, where $\beta = (\beta_1, \beta_2)$ and $\hat{\beta}_j$'s are LSE's under the reduced models.

Let $H = (H_1 \ 0)$ be $n \times p$.

Note that

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

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

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

which is not 0 by Theorem 3.10.

This implies that some LSE $l^\tau \hat{\beta}$ is not a BLUE of $l^\tau \beta$ but $l^\tau \hat{\beta}$ is the BLUE of $l^\tau \beta$ if $l \in \mathcal{R}(H)$.

Finally, we consider model (1) with $\text{Var}(\varepsilon)$ being a diagonal matrix whose i th diagonal element is σ_i^2 , i.e., ε_i 's are uncorrelated but have unequal variances.

A straightforward calculation shows that condition (e) in Theorem 3.10 holds if and only if, for all $i \neq j$, $\sigma_i^2 \neq \sigma_j^2$ only when $h_{ij} = 0$, where h_{ij} is the (i, j) th element of the projection matrix $Z(Z^\tau Z)^- Z^\tau$.

Thus, an LSE is not a BLUE in general, although it is still unbiased for estimable $l^\tau \beta$.

Suppose that the unequal variances of ε_i 's are caused by some small perturbations, i.e., $\varepsilon_i = e_i + u_i$, where $\text{Var}(e_i) = \sigma^2$, $\text{Var}(u_i) = \delta_i$, and e_i and u_i are independent so that $\sigma_i^2 = \sigma^2 + \delta_i$.

$$\text{Var}(l^\tau \hat{\beta}) = l^\tau (Z^\tau Z)^- \sum_{i=1}^n \sigma_i^2 Z_i Z_i^\tau (Z^\tau Z)^- l.$$

If $\delta_i = 0$ for all i (no perturbations), then assumption A2 holds and $l^\tau \hat{\beta}$ is the BLUE of any estimable $l^\tau \beta$ with $\text{Var}(l^\tau \hat{\beta}) = \sigma^2 l^\tau (Z^\tau Z)^- l$.

Suppose that $0 < \delta_i \leq \sigma^2 \delta$. Then

$$\text{Var}(l^\tau \hat{\beta}) \leq (1 + \delta) \sigma^2 l^\tau (Z^\tau Z)^- l.$$

This indicates that the LSE is robust in the sense that its variance increases slightly when there is a slight violation of the equal variance assumption (small δ).