L01 Least square estimators

1. LSE(β)

   (1) LSE of β
   In linear model $y = X\beta + e$ with random error $e = y - X\beta \sim (0, \sigma^2 I_n)$,
   $\|e\|^2 = \sum_{i=1}^n e_i^2 = \|y - X\beta\|^2 = Q(\beta)$.

$$\begin{aligned} & \widehat{\beta} \text{ is a least square estimator (LSE) for } \beta \Longleftrightarrow Q(\widehat{\beta}) \le Q(\beta)\ \forall \beta \in R^p \\ \Longleftrightarrow\ & \|y - X\widehat{\beta}\|^2 \le \|y - X\beta\|^2\ \forall\beta \Longleftrightarrow X\widehat{\beta} = \pi(y|\mathcal{R}(X)) \\ \Longleftrightarrow\ & X\widehat{\beta} = XX^+y \text{ (Projection equation)} \\ \Longleftrightarrow\ & X'X\widehat{\beta} = X'y \text{ (Normal equation)} \end{aligned}$$

   (2) LSE(β)
   Let LSE(β) be the collection of all LSEs of β. Then LSE(β) $= X^+y + \mathcal{N}(X)$.
   Proof $\widehat{\beta} \in \text{LSE}(\beta) \iff X\widehat{\beta} = XX^+y \iff X(\widehat{\beta} - X^+y) = 0$
   $\iff \widehat{\beta} - X^+y \in \mathcal{N}(X) \iff \widehat{\beta} \in X^+y + \mathcal{N}(X)$.

   (3) Minimum norm LSE for β
   X^+y is the minimum norm LSE for β.
   Proof $X^+y = X^+y + 0 \in X^+y + \mathcal{N}(X) = \text{LSE}(\beta)$
   For $\widehat{\beta} \in \text{LSE}(\beta) = X^+y + \mathcal{N}(X)$, $\widehat{\beta} = X^+y + Z$ where $Z \in \mathcal{N}(X)$.
   But $\langle X^+y, Z\rangle = Z'X^+y = Z'X^+XX^+y = (XZ)'(X^+)'X^+y = 0$.
   So $\|\widehat{\beta}\|^2 = \|X^+y\|^2 + \|Z\|^2 \ge \|X^+y\|^2$.

Ex1: A special case
$X \in R^{n\times p}$ has full column rank $\iff \dim[\mathcal{N}(X)] = p - \text{rank}(X) = 0 \iff \mathcal{N}(X) = \{0\}$
$\iff \text{LSE}(\beta) = \{X^+y\}$.

Ex2: $(X'X)^-X'y \subset \text{LSE}(\beta)$.
$\widehat{\beta} \in (X'X)^-X'y \Longrightarrow X\widehat{\beta} = X(X'X)^-X'y = XX^+y \Longrightarrow \widehat{\beta} \in \text{LSE}(\beta)$.

Comment: LSE(β) $\subset (X'X)^-X'y$ is not true, i.e., $X^+y + \mathcal{N}(X) \subset (X'X)^-X'y$ is not true.
We only need to see the inclusion is not true when $y = 0$, i.e., N$(X) \subset \{0\}$ is not true.
Hence Theorem 5.1.1 on p150 is false.

2. LUE($A\beta$)

   (1) LUE($A\beta$)
   $\widehat{\gamma}$ is a linear unbiased estimator (LUE) for $A\beta \overset{def}{\Longleftrightarrow} \widehat{\gamma} = Ly$ and $E(Ly) = A\beta\ \forall\beta$
   $\iff \widehat{\gamma} = Ly,\ LX = A$.
   Let LUE($A\beta$) be the collection of all LUEs for $A\beta$. Then LUE($A\beta$) $= \{Ly : LX = A\}$.

   (2) Estimable $A\beta$
   $A\beta$ is estimable $\overset{def}{\Longleftrightarrow} \text{LUE}(A\beta) \ne \emptyset \iff A = LX$ for some L
   $\overset{*}{\Longleftrightarrow} A \cdot \text{LSE}(\beta) = \{AX^+y\}$

   $\overset{*}{\Rightarrow}$**:** $A = LX \Longrightarrow A\cdot\text{LSE}(\beta) = AX^+y + A\mathcal{N}(X) = AX^+y + LX\mathcal{N}(X) = \{AX^+y\}$.
   $\overset{*}{\Leftarrow}$**:** $A\cdot\text{LSE}(\beta) = AX^+y \Longrightarrow A\mathcal{N}(X) = \{0\} \Longrightarrow A\mathcal{R}(I - X^+X) = \mathcal{R}[A(I - X^+X)] = \{0\}$
   So $0 = \dim[\mathcal{R}(A - AX^+X)] = \text{rank}(A - AX^+X) \Longrightarrow A = AX^+X$.
   Hence $A = LX$ for some L.

(3) LSE($A\beta$)
For estimable $A\beta$, define the LSE of $A\beta$ as $\text{LSE}(A\beta) = A \cdot \text{LSE}(\beta) = AX^+y$.
Then $\text{LSE}(A\beta) \in \text{LUE}(A\beta)$.

Proof $A\beta$ is estimable. Then $LX = A$ for some L.
So $\text{LSE}(A\beta) = AX^+y = LXX^+y$ with $LXX^+X = LX = A$.
Thus $AX^+y \in \text{LUE}(A\beta)$.

Ex3: β is estimable $\Longleftrightarrow$ $A\beta$ is estimable for all A.

$\Rightarrow$: If β is estimable, then $I_p = LX$ for some L. So X has left-inverse X^L.
Thus for all $A \in R^{q\times p}$, $A = AX^LX = LX$ for some L. Hence $A\beta$ is estimable.

$\Leftarrow$: $A\beta$ is estimable for all A. Then $\beta = I_p\beta$ is estimable.

3. BLUE of estimable $A\beta$

(1) BLUE
For $\widehat{\gamma}$, an estimator for γ, $r_{\widehat{\gamma}} = E[(\widehat{\gamma}-\gamma)(\widehat{\gamma}-\gamma)']$ is a risk function.
If $\widehat{\gamma}$ is an unbiased estimator for γ, then $r_{\widehat{\gamma}} = \text{Cov}(\widehat{\gamma})$.
$\widehat{\gamma}$ is a best linear unbiased estimator (BLUE) for γ by the above risk
if $\widehat{\gamma} \in \text{LUE}(\gamma)$ and $\text{Cov}(\widehat{\gamma}) \le \text{Cov}(\widehat{\eta})$ for all $\widehat{\eta} \in \text{LUE}(\gamma)$. Here the inequality means that $\text{Cov}(\widehat{\eta})) - \text{Cov}(\widehat{\gamma})$ is non-negative definite.

(2) Theorem
If $A\beta$ is estimable, then $\text{LUE}(A\beta) = AX^+y$ is a BLUE for $A\beta$.

Proof $AX^+y \in \text{LUE}(A\beta)$. If $Ly \in \text{LUE}(A\beta)$, then $A = LX$.
So $\text{Cov}(Ly) - \text{Cov}(AX^+y) = \sigma^2L(I - XX^+)L' = \text{Cov}(L(I - XX^+)y) \ge 0$.

(3) Uniqueness
If Ly is also a BLUE for $A\beta$, then $Ly = AX^+y$.

Proof If Ly is also BLUE for $A\beta$, then $\text{Cov}(Ly) - \text{Cov}(AX^+y) = \sigma^2L(I - XX^+)L' \ge 0$ and $\text{Cov}(AX^+y) - \text{Cov}(Ly) = -\sigma^2L(I - XX^+)L' \ge 0$. Hence $\sigma^2L(I - XX^+)L' = 0$.
Thus $L = LXX^+ = AX^+$. Therefore $Ly = AX^+y$.