L02: BLUE and BLUP

1. A theorem and preliminaries

   (1) Theorem

   Suppose $\begin{pmatrix} y \\ y_* \end{pmatrix} \sim \left(\begin{pmatrix} X \\ X_* \end{pmatrix} \mu, \, \sigma^2 \begin{pmatrix} \Sigma & C \\ C' & V \end{pmatrix} \right)$ where $E(y_*) = X_* \mu = \theta$.

   (i) By is BLUE (best linear unbiased estimator) for θ with respect to MSCPE risk
   $\Longleftrightarrow BX = X_*$ and $B\Sigma(I - XX^+) = 0$

   (ii) By is BLUP (best linear unbiased predictor) for y_* with respect to MSCPE risk
   $\Longleftrightarrow BX = X_*$ and $(B\Sigma - C')(I - XX^+) = 0$.

   (2) X^+

   X^+ is the existent and unique matrix satisfying the four conditions $XX^+X = X$, $X^+XX^+ = X^+$, $(XX^+)' = XX^+$ and $(X^+X)' = X^+X$.
   This X^+ is called Moore-Penrose inverse of X.
   $I - XX^+$ and $I - X^+X$ are both symmetric and idempotent.

   (3) $AB = 0 \Longleftrightarrow A \in \{H(I - BB^+) : H\}$. Here H and A are of the same dimensions.

   ⇒: If $AB = 0$, then $ABB^+ = 0$.
   So $A = A(I - BB^+ + BB^+) = A(I - BB^+) \in \{H(I - BB^+) : H\}$.

   ⇐: If $A \in \{H(I - BB^+) : H\}$, then $A = H(I - BB^+)$ for some H.
   So $AB = H(I - BB^+)B = H(B - B) = 0$.

   (4) Definite and semi-definite matrices

   For $A' = A$, $\quad A > 0 \Longleftrightarrow$ all eigenvalues of A are > 0;
   $\qquad\qquad\quad A < 0 \Longleftrightarrow$ all eighenvalues of A are < 0.
   If $A \geq 0$, then $BAB' \geq 0$ since $x'BAB'x = y'Ay \geq 0$ for all x where $y = Bx$.
   If $A \leq 0$, then $BAB' \leq 0$ since $x'BAB' = y'Ay \leq 0$ for all x where $y = Bx$.

2. Proof the iff conditions for BLUE

   (1) $\Leftarrow$ (The condition is sufficient)

   $BX = X_* \Longrightarrow E(By) = BX\mu = X_*\mu = \theta \Longrightarrow By \in \text{LUE}(\theta)$.
   If $Ty \in \text{LUE}(\theta)$, then $TX\mu = E(Ty) = \theta = X_*\mu$ for all μ. So $TX = X_*$.
   Now $TX = X_* = BX \Longrightarrow (T - B)X = 0 \Longrightarrow T - B \in \{H(I - XX^+) : H\}$. Thus

   $$T \in B + \{H(I - XX^+) : H\}.$$

   With given condition $B\Sigma(I - XX^+) = 0$, because $\Sigma > 0$,

   $$\begin{aligned} \text{Cov}(Ty) - \text{Cov}(By) &= \sigma^2(T\Sigma T' - B\Sigma B') \\ &= \sigma^2 \left\{[B + H(I - XX^+)]\Sigma[B + H(I - XX^+)]' - B\Sigma B'\right\} \\ &= \sigma^2 H(I - XX^+)\Sigma(I - XX^+)H' \\ &= \sigma^2[H(I - XX^+)]\Sigma[H(I - XX^+)]' \geq 0. \end{aligned}$$

   Thus By is a minimum variance-covariance matrix LUE for $\theta = X_*\mu$.
   Therefore it is BLUE for θ with respect to MSCPE risk.

(2) $\Rightarrow$ (The condition is necessary)
Suppose By is BLUE for θ. We show $BX = X_*$ and $B\Sigma(I - XX^+) = 0$.
$By \in \text{LUE}(\theta)$. So $X_*\mu = \theta = E(By) = BX\mu$ for all μ. Thus $BX = X_*$.
If $Ty \in \text{LUE}(\theta)$, by the proof in (1), $T = B + H(I - XX^+)$.
From $0 \le \text{Cov}(Ty) - \text{Cov}(By)$ for all $Ty \in \text{LUE}(\theta)$,

$$B\Sigma(I - XX^+)H' + H(I - XX^+)\Sigma B' + H(I - XX^+)\Sigma(I - XX^+)H' \ge 0 \text{ for all } H.$$

For $\Sigma > 0$, in the EVD $\Sigma = P\Lambda P'$ $\Lambda = \text{diag}(\lambda_1, ..., \lambda_n)$ with $\lambda_i > 0\ \forall\, i$.
Let $0 < \lambda < \frac{2}{\max(\lambda_1,..,\lambda_n)}$ and $H = -\lambda B\Sigma(I - XX^+)$. The displayed inequality becomes

$$\begin{aligned}&[B\Sigma(I - XX^+)](\lambda^2\Sigma - 2\lambda I)[B\Sigma(I - XX^+)]'\\ &\qquad = [B\Sigma(I - XX^+)]P\Gamma P'[B\Sigma(I - XX^+)]' \ge 0\end{aligned}$$

where $\Gamma = \text{diag}(\lambda_1\lambda^2 - 2\lambda, ..., \lambda_n\lambda^2 - 2\lambda) < 0$ since $\lambda_i\lambda^2 - 2\lambda = \lambda_i\lambda\left(\lambda - \frac{2}{\lambda_i}\right) < 0\ \forall\, i$.
Suppose $B\Sigma(I - XX^+) \ne 0$. WLOG assume the first column of $[B\Sigma(I - XX^+)]'$, $\alpha = [B\Sigma(I - XX^+)]'e_1 \ne 0$. Then $\beta = P'\alpha \ne 0$. Now

$$e_1'[B\Sigma(I - XX^+)]P\Gamma P'[B\Sigma(I - XX^+)]'e_1 = \alpha' P\Gamma P'\alpha = \beta'\Gamma\beta < 0.$$

This contradiction implies that $B\Sigma(I - XX^+) \ne 0$ is false. Thus $B\Sigma(I - XX^+) = 0$.

3. Proof iff condition for BLUP

(1) $\Leftarrow$ (The condition is sufficient)
$BX = X_* \Longrightarrow By \in \text{LUE}(E(y_*)) \Longrightarrow By \in \text{LUP}(y_*)$.
If $Ty \in \text{LUP}(y_*)$, then $Ty \in \text{LUE}(E(y_*))$. By the proof in 1, $T \in B + \{H(I - XX^+) :\ H\}$.
With $D = (B\Sigma - C')(I - XX^+)$,

$$\begin{aligned}&\text{Cov}(Ty - y_*) - \text{Cov}(By - y_*)\\ &\qquad = \sigma^2\left\{[H(I - XX^+)]\Sigma[H(I - XX^+)]' + HD' + DH'\right\}.\end{aligned}$$

$D = 0 \Longrightarrow \text{Cov}(Ty - y_*) - \text{Cov}(By - y_*) = [H(I - XX^+)]\Sigma[H(I - XX^+)]' \ge 0$.
So By is BLUP for y_* with respect to MSCPE risk.

(2) $\Rightarrow$ (The condition is necessary)
By is BLUP for $y_* \Longrightarrow By \in \text{LUP}(y_*) = \text{LUE}(E(y_*)) \Longrightarrow BX = X_*$.
If $Ty \in LUP(y_*)$, then $Ty \in \text{LUE}(E(y_*))$. So $T \in B + \{H(I - XX^+) :\ H\}$.
By is BLUP for $y_* \Longrightarrow \text{Cov}(Ty - y_*) - \text{Cov}(By - y_*) \ge 0$,

$$H(I - XX^+)\Sigma(I - XX^+)H' + HD' + DH' \ge 0 \text{ for all } H$$

where $D = (B\Sigma - C')(I - XX^+)$. Select λ as in (2) of 2 and let $H = -\lambda D$. Then

$$D(\lambda^2\Sigma - 2\lambda I)D' \ge 0$$

Suppose $D \ne 0$. WOLG assume the first column of D', $\alpha = D'e_1 \ne 0$.
By the same argument as in (2) of 2, $e_1'D(\lambda^2\Sigma - 2\lambda)D'e_1 = \alpha'(\lambda^2\Sigma - 2\lambda I)\alpha < 0$.
The contradiction shows that $D = 0$, i.e., $(B\Sigma - C')(I - XX^+) = 0$.

L03: Consistent estimators

1. Sample mean

   (1) $\overline{y} \in \text{LUE}(\mu)$
   Sample mean $\overline{y}$ is a LUE for the population mean μ, i.e., $\overline{y} \in \text{LUE}(\mu)$. But is it the best one in LUE(μ) or even in UE(μ) by MSCPE risk?

   (2) Two sufficient conditions for $\widehat{\theta}$ to be the best one in UE(θ) wrt MSCPE risk.
   (i) $E(\widehat{\theta}) = \theta$ and $\widehat{\theta}$ is a function of a sufficient and complete statistic S.
   (ii) $E(\widehat{\theta}) = \theta$ and $\text{Cov}(\widehat{\theta}) = CRLB(\theta)$.

   Caution: If a sufficient condition does not hold, that does not mean the conclusion does not hold since a sufficient condition may not be a necessary one. See 1 in HW01.

   (3) Sufficient and necessary conditions
   In L02 we learned a sufficient and necessary condition for By to be the BLUE of θ, and for By to be the BLUP for y_*. We show that sample mean is BLUE for population mean.

   Write $\begin{pmatrix} y \\ y_* \end{pmatrix} \sim \left(\begin{pmatrix} X \\ X_* \end{pmatrix} \mu, \begin{pmatrix} \Sigma & C \\ C' & V \end{pmatrix} \right)$ where $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, $y_1, ..., y_n$ is a random sample from a population with parameters (μ, V). Let $\overline{y} = By$ be the sample mean.
   By HW01, $X = 1_n \otimes I_k$, $X_* = I_k$, $\Sigma = I_n \otimes V$, $C = 0$ and $B = \frac{1_n' \otimes I_k}{n}$.

   Now $BX = \frac{1_n' \otimes I_k}{n}(1_n \otimes I_k) = \frac{n \otimes I_k}{n} = I_k = X_*$.

   $$\begin{aligned} B\Sigma(I - XX^+) &= \tfrac{1_n' \otimes I_k}{n}(I_n \otimes V)\left[I - (1_n \otimes I_k)(1_n \otimes I_k)^+\right] \\ &= \tfrac{1_n' \otimes V}{n}\left[I - (1_n \otimes I_k)(1_n^+ \otimes I_k^+)\right] = \tfrac{1_n' \otimes V}{n}\left[I - 1_n 1_n^+ \otimes I_k\right] \\ &= \tfrac{1_n' \otimes V}{n} - \tfrac{1_n' \otimes V}{n} = 0. \end{aligned}$$

   So $\overline{y}$ is BLUE for μ by MSCPE risk.
   Comment: $\overline{y}$ is also BLUP for y_* since $C = 0$.

2. Strongly consistent estimators

   (1) Asymptotically unbiased estimator
   If $E(\widehat{\theta}_n) = \theta$, then $\widehat{\theta}_n$ is an UE for θ — from statical point of view.
   If $E(\widehat{\theta}_n) \neq \theta$, but $\lim_n E(\widehat{\theta}_n) = \theta$, then $\widehat{\theta}_n$ is an asymptotically UE for θ –from dynamic point of view.

   (2) Strongly consistent estimator
   To define more asymptotical properties of $\widehat{\theta}_n$ we need the convergence of random variable/vectors.

   X_n converges to X almost sure, or with probability 1 denoted as $X_n \xrightarrow{a.s.} X$ or $X_n \xrightarrow{wp1} X$ if $P(X_n \longrightarrow X) = 1$.
   If $\widehat{\theta}_n \xrightarrow{a.s.} \theta$, then $\widehat{\theta}_n$ is called a strongly consistent estimator for θ

   (3) Strong law of large numbers (SLLN)
   If $X_1, ..., X_n$ are iid with (μ, Σ), then $\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} \mu$.
   Thus sample mean is always a strong consistent estimator for population mean.

(4) Properties

$$X_n \xrightarrow{a.s.} X \iff g(X_n) \xrightarrow{a.s.} g(X) \text{ for all continuous } g(\cdot)$$
$$X_n \xrightarrow{a.s.} X \iff X_{n_k} \xrightarrow{a.s.} X \text{ for all subsequence } X_{n_k}.$$
$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{a.s.} \begin{pmatrix} X \\ Y \end{pmatrix} \iff X_n \xrightarrow{a.s.} X \text{ and } Y_n \xrightarrow{a.s.} Y.$$

3. Consistent estimators

(1) Consistent estimators

$X_n \xrightarrow{p} X$ (X_n converges to X in probability) if $\lim_n P(\|X_n - X\| < \epsilon) = 1$ for all $\epsilon > 0$.

$X_n \xrightarrow{p} X \iff X_n - X \xrightarrow{p} 0 \overset{def}{\iff} X_n - X = o_p(1)$.

If $\widehat{\theta}_n \xrightarrow{p} \theta$, then $\widehat{\theta}_n$ is called a consistent estimator for θ.

(2) Relations

$X_n \xrightarrow{a.s.} X \Longrightarrow X_n \xrightarrow{p} X \qquad X_n \xrightarrow{p} X \Longrightarrow \exists\, X_{n_k}$ such that $X_{n_k} \xrightarrow{a.s.} X$.

(3) Weak law of large numbers (WLLN)

If $X_1, ..., X_n$ are iid with (μ, Σ), then $\frac{X_1+\cdots+X_n}{n} \xrightarrow{p} \mu$

(4) Properties

$$X_n \xrightarrow{p} X \iff g(X_n) \xrightarrow{p} g(X) \text{ for all continuous } g(\cdot)$$
$$X_n \xrightarrow{p} X \iff X_{n_k} \xrightarrow{p} X \text{ for all subsequence } X_{n_k}.$$
$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{p} \begin{pmatrix} X \\ Y \end{pmatrix} \iff X_n \xrightarrow{p} X \text{ and } Y_n \xrightarrow{p} Y.$$

Ex: Let $\overline{y}_n$ be sample mean and μ be the population mean. By SLLN or WLLN and their properties $\overline{y} \xrightarrow{p} \mu \Longrightarrow (\overline{y})^2 \xrightarrow{p} \mu^2$, i.e., $(\overline{y}_n)^2$ is a consistent estimator for μ^2.

(5) Sequence bounded in probability

X_n is bounded in probability denoted as $X_n = O_p(1)$ if $\forall\, \epsilon > 0, \exists\, M_\epsilon > 0$ and $N_\epsilon > 0$ such that $P(\|X_n\| < M_\epsilon) > 1 - \epsilon$ for all $n > N_\epsilon$.

If $\sqrt{n}(\widehat{\theta}_n - \theta) = O_p(1)$, then

$$\widehat{\theta}_n - \theta = \frac{1}{\sqrt{n}}[\sqrt{n}(\widehat{\theta}_n - \theta)] = o_p(1)O_p(1) = o_p(1) \Longrightarrow \widehat{\theta}_n - \theta \xrightarrow{p} 0 \iff \widehat{\theta}_n \xrightarrow{p} \theta,$$

i.e., $\widehat{\theta}_n$ is a consistent estimator for θ.

The above is true when θ is population median and $\widehat{\theta}_n$ is sample median.

So sample median is a consistent estimator for population median.