L02: Linear unbiased estimators

1. Estimability of parameters

   (1) Linear unbiased estimators and estimable parameters
   $\widehat{\eta}$ is a linear unbiased estimator (LUE) for η if
   (i) $\widehat{\eta}$ is a linear function of data vector Y, i.e., $\widehat{\eta} = LY$ for some L.
   (ii) $\widehat{\eta}$ is an unbiased estimator for η, i.e., $E(\widehat{\eta}) \equiv \eta$.
   Parameter vector η is estimable if it has a linear unbiased estimator For parameter η, let LUE(η) be the collection of all its LUEs. Then

   $$\eta \text{ is estimable } \Longleftrightarrow \text{LUE}(\eta) \neq \emptyset.$$

   Comments: For linear model $Y = X\beta + \epsilon$ with $E(\epsilon) = 0$, the parameters of interests are linear functions of β, $H\beta$. Among them we are interested in the estimable ones. For estimable parameters we will try to find their linear unbiased estimators.

   Ex1: For the liner model, Y is a LUE for $X\beta$. So $X\beta$ is estimable, and $Y \in \text{LUE}(X\beta)$.

   (2) Estimable $H\beta$
   For linear model $Y = X\beta + \epsilon$ with $E(\epsilon) = 0$, $H\beta$ is estimable if and only if it is a linear function of $X\beta$, i.e.,

   $$H\beta \text{ is estimable } \Longleftrightarrow H = LX \text{ for some } L$$

   Proof. $H\beta$ is estimable $\Longleftrightarrow \exists\, LY$ such that $E(LY) \equiv H\beta$
   $\Longleftrightarrow \exists\, LY$ such that $LX\beta = X\beta$ for all β
   $\Longleftrightarrow \exists\, L$ such that $LX = H$.

   (3) LUE($H\beta$)
   $\text{LUE}(H\beta) = \{LY : H = LX\}$.

   Proof. $\widehat{\eta} \in \text{LUE}(H\beta) \Longleftrightarrow \widehat{\eta} = LY$ and $E(LY) \equiv H\beta \Longleftrightarrow \widehat{\eta} = LY$ and $LX = H$
   $\Longleftrightarrow \widehat{\eta} \in \{LY : LX = H\}$.

2. Relation of estimability and GLSE

   (1) Sufficient and necessary conditions
   For linear model $Y = X\beta + \epsilon$ with $E(\epsilon) = 0$, the followings are equivalent.
   (i) $H\beta$ is estimable (ii) $H\mathcal{N}(X) = \{0\}$ (iii) $H[\text{GLSE}_U(\beta)] = H\left(U^{1/2}X\right)^{+}U^{1/2}Y$.

   Proof. (i)$\Longleftrightarrow$ (ii)
   $\Rightarrow$: If $H\beta$ is estimable, then $H = LX$ for some L. So $H\mathcal{N}(X) = LX\mathcal{N}(X) = \{0\}$.
   $\Leftarrow$: If $H\mathcal{N}(X) = \{0\}$, then $\mathcal{N}(X) \subset \mathcal{N}(H)$. So $\mathcal{N}^{\perp}(H) \subset \mathcal{N}^{\perp}(X)$.
   Thus $\mathcal{R}(H') \subset \mathcal{R}(X')$. Hence $H' = X'L'$ for some L, i.e., $H = LX$ for some L.
   Therefore $H\beta$ is estimable.
   (ii)$\Longleftrightarrow$(iii)
   Note that $\text{GLSE}_U(\beta) = \left(U^{1/2}X\right)^{+}U^{1/2} + \mathcal{N}(X)$.
   So, $H[\text{GLSE}_U(\beta)] = H\left(U^{1/2}X\right)^{+}U^{1/2}Y + H\mathcal{N}(X)$. (ii)$\Longleftrightarrow$(iii) follows.

   Comments: Here $\widehat{\beta} = \left(U^{1/2}X\right)^{+}U^{1/2}Y$ is the minimum norm GLSE.
   When showing the equivalency, U could be any qualified matrices.
   Thus (iii) could be $H[\text{OLSE}(\beta)] = HX^{+}Y$.

(2) Let $\widehat{\beta} = \left(U^{1/2}X\right)^{+} U^{1/2}Y$, the minimum norm GLSE.
If $H\beta$ is estimable, then $H\widehat{\beta} \in \text{LUE}(H\beta)$.

Proof. $H\widehat{\beta} = H\left(U^{1/2}X\right)^{+} U^{1/2}Y$ is a linear function of Y.
When $H\beta$ is estimable, $H = LX$ for some L. So

$$\begin{aligned} E(H\widehat{\beta}\,) &= LX\left(U^{1/2}X\right)^{+} U^{1/2}X\beta = LU^{-1/2}\left(U^{1/2}X\right)\left(U^{1/2}X\right)^{+}\left(U^{-1/2}X\right)\beta \\ &= LU^{-1/2}\left(U^{1/2}X\right)\beta = LX\beta = H\beta. \end{aligned}$$

Thus $H\widehat{\beta} \in \text{LUE}(H\beta)$.

Ex2: $\left\{H\left(U^{1/2}X\right)^{+} U^{1/2}Y : U > 0\right\} \subset \text{LUE}(H\beta)$.

3. LUE($H\beta$)

(1) Expression I
For estimable $H\beta \in R^q$, if $L_0Y \in \text{LUE}(H\beta)$, then

$$LUE(H\beta) = L_0Y + \mathcal{N}(I_q,\, X)Y$$

Proof. $L_0Y \in \text{LUE}(H\beta) \Longleftrightarrow L_0X = H$.

$$\begin{aligned} LY \in \text{LUE}(H\beta) &\Longleftrightarrow LX = H = L_0X \Longleftrightarrow I_q(L - L_0)X = 0 \\ &\Longleftrightarrow L - L_0 \in \mathcal{N}(I_q,\, X) \Longleftrightarrow L \in L_0 + \mathcal{N}(I_q,\, X) \\ &\Longleftrightarrow LY \in L_0Y + \mathcal{N}(I_q,\, X)Y. \end{aligned}$$

Comment: $AXB = 0 \Longleftrightarrow X \in \mathcal{N}(A,\, B)$.

(2) Expression II
For estimable $H\beta \in R^q$,

$$\text{LUE}(H\beta) = H\left(U^{1/2}X\right)^{+} U^{1/2}Y + \mathcal{N}(I_q,\, X)Y.$$

Proof. By Ex2, $L_0Y = H\left(U^{1/2}X\right)^{+} U^{1/2}Y \in \text{LUE}(H\beta)$.
Conclusion then follows from (1).

Comment: In this expression, $U \in R^{n\times n}$ could be any qualified matrices.

Ex3: With $U = I$, $\text{LUE}(H\beta) = HX^{+}Y + \mathcal{N}(I_q,\, X)Y = [HX^{+} + \mathcal{N}(I_q,\, X)]Y$.

(3) Comments

(i) In $\text{LUE}(H\beta) = (L_0 + \mathcal{N}(I_q,\, X))Y$, $\mathcal{M} = L_0 + \mathcal{N}(I_q,\, X)$ is an affine set in $R^{q\times n}$ just like $\text{OLSE}(\beta) = X^{+}Y + \mathcal{N}(X)$ and $\text{GLSE}_U(\beta) = \left(U^{1/2}X\right)^{+} U^{1/2}Y + \mathcal{N}(X)$ are affine sets in R^p.

(ii) The role of L_0 in the expression of $\mathcal{M} = L_0 + \mathcal{N}(I_q,\, X)$ can be replaced by any $L_1 \in \mathcal{M}$, i.e., $L_1 \in \mathcal{M} \Longrightarrow L_1 + \mathcal{N}(I_q,\, X) = L_2 + \mathcal{N}(I_q,\, X)$.

(iii) It is possible to find L_3 such that $\mathcal{M} = L_3 + \mathcal{N}(I_q,\, X)$ and $L_3 \perp \mathcal{N}(I_q,\, X)$. This L_3 is the minimum norm matrix in $\mathcal{M}$. However to do it one must specify the inner product in $R^{q\times n}$.

(iv) With respect to the Frobenius inner product in $R^{q\times n}$, in Ex3

$$\mathcal{M} = HX^{+} + \mathcal{N}(I_q,\, X) \text{ and } HX^{+} \perp \mathcal{N}(I_q,\, X).$$

L03 Linear unbiased predictors

1. Concepts of linear unbiased predictors

   (1) Predictor
   Linear model on data $Y \in R^n$: $Y = X\beta + \epsilon$ where $E(\epsilon) = 0$
   Future response $Y_f \in R^q$: $Y_f = H\beta + \epsilon_f$ where $E(\epsilon_f) = 0$
   Y and Y_f share the same β. Both X and H are known. Y_f needs to be predicted by $\widehat{Y}_f$, a vector valued function of Y. In such a case $\widehat{Y}_f$ is called a predictor for Y_f.

   (2) Linear unbiased predictor
   $\widehat{Y}_f$ is a linear predictor (LUP) for Y_f if
   (i) $\widehat{Y}_f$ is a linear predictor, i.e., $\widehat{Y}_f = LY$ for some L.
   (ii) $\widehat{Y}_f$ is an unbiased predictor, i.e., $E\left(\widehat{Y}_f - Y_f\right) \equiv 0$.

   (3) Predictability
   Y_f may or may not have a LUP. We say that Y_f is predictable if it does have a LUP. Let $\text{LUP}(Y_f)$ be the collection of all LUP for Y_f. Then

$$Y_f \text{ is predictable } \Longleftrightarrow \text{LUP}(Y_f) \neq \emptyset$$

2. Predictability and the collection of all LUPs

   (1) Relations
   With $E(Y_f) = H\beta$,
   (i) $\text{LUP}(Y_f) = \text{LUE}(H\beta)$.
   (ii) Y_f is predictable $\Longleftrightarrow E\beta$ is estimable

   Proof. (i)
$$\begin{aligned} LY \in \text{LUP}(Y_f) &\Longleftrightarrow E(LY - Y_f) \equiv 0 \Longleftrightarrow E(LY) \equiv H\beta \\ &\Longleftrightarrow LY \in \text{LUE}(H\beta) \end{aligned}$$

   (ii)
$$\begin{aligned} Y_f \text{ is predictable} &\Longleftrightarrow \exists\, LY \text{ such that } LY \in \text{LU}(Y_f) \\ &\Longleftrightarrow \exists\, LY \text{ such that } LY \in \text{LUE}(H\beta) \\ &\Longleftrightarrow H\beta \text{ is estimable} \end{aligned}$$

   (2) Sufficient and necessary conditions for predictability
   Under $E(Y_f) = H\beta$,

$$\begin{aligned} &Y_f \text{ is predictable} \Longleftrightarrow \text{LUP}(Y_f) \neq \emptyset \Longleftrightarrow \text{LUE}(H\beta) \neq \emptyset \Longleftrightarrow H\beta \text{ is estimable} \\ &\Longleftrightarrow H = LX \text{ for some } L \Longleftrightarrow H\mathcal{N}(X) = \{0\} \\ &\Longleftrightarrow H\left[\text{GLES}_U(\beta)\right] = H\left(U^{1/2}X\right)^+ U^{1/2}Y. \end{aligned}$$

   (3) $\text{LUP}(Y_f)$.
   Suppose $L_0Y \in \text{LUP}(Y_f) = \text{LUE}(H\beta)$. Then

$$\begin{aligned} \text{LUP}(Y_f) = \text{LUE}(H\beta) &= L_0Y + \mathcal{N}(I_q,\, X)Y \\ &= H\left(U^{1/2}X\right)^+ U^{1/2}Y + \mathcal{N}(I_q,\, X)Y. \end{aligned}$$

3. Matrix valued risk functions

(1) Matrix valued risk function
When using statistic $\widehat{u} \in R^q$ to estimate v (v is a parameter vector), or to predict v (v is a random vector), the loss can be measured by $(\widehat{u}-v)(\widehat{u}-v)' \in R^{q\times q}$ with risk

$$r(\widehat{u},\, v) = E[(\widehat{u}-v)(\widehat{u}-v)'] \in R^{q\times q}.$$

The risk matrix $r(\widehat{u},\, v)$ is a non-negative definite matrix, denoted by $r(\widehat{u},\, v) \geq 0$.

(2) Usage
$\widehat{u}$ and $\widetilde{u}$ are two estimators/predictors for v.

$$\widehat{u} \text{ dominates } \widetilde{u} \overset{def}{\Longleftrightarrow} r(\widehat{u},\, v) \leq r(\widetilde{u},\, v) \text{ at all parameter points.}$$

Here $A \leq B$ means $A - B \leq 0$, i.e., $A - B$ is a non-positive definite matrix, or equivalently $B - A \geq 0$, i.e., $B - A$ is a non-negative definite matrix.

In a class of estimators/predictors, $\widehat{u}$ is inadmissable if it is dominated by another estimator/predictor. Otherwise it is admissable.

In a class of estimators/predictor, $\widehat{u}$ is the best estimator/predictor if it dominates all other estimators/predictors.

(3) Bias and risk
When estimating/predicting v by $\widehat{u}$, $b = E(\widehat{u}-v) \in R^q$ is the bias.

$$\begin{aligned} r(\widehat{u},\, v) &= E[(\widehat{u}-v)(\widehat{u}-v)'] = E\{[(\widehat{u}-v-b)+b][(\widehat{u}-v-b)+b]'\} \\ &= \mathrm{Cov}(\widehat{u}-v) + bb'. \end{aligned}$$

$\widehat{u}$ is unbiased estimator/predictor if its bias $b = 0$. For unbiased estimator $\widehat{u}$,

$$r(\widehat{u},\, v) = \mathrm{Cov}(\widehat{u}-v) = \mathrm{Cov}(\widehat{u}).$$

For unbiased predictor $\widehat{u}$,

$$r(\widehat{u},\, v) = \mathrm{Cov}(\widehat{u}-v).$$

Ex1: Suppose $\widehat{u} \in R^q$ dominates $\widetilde{u} \in R^q$ when estimating/predicting $v \in R^q$.
Most likely, one would use $A\widehat{u}$ and $A\widetilde{u}$ to estimate/predict Av.

$$\begin{aligned} & r(\widehat{u},\, v) \leq r(\widetilde{u},\, v) \Longleftrightarrow E[(\widehat{u}-v)(\widehat{u}-v)'] \leq E[(\widetilde{u}-v)(\widetilde{u}-v)'] \\ \Longrightarrow\; & A\,E[(\widehat{u}-v)(\widehat{u}-v)']A' \leq A\,E[(\widetilde{u}-v)(\widetilde{u}-v)']A' \\ \Longleftrightarrow\; & E[(A\widehat{u}-Av)(A\widehat{u}-Av)'] \leq E[(A\widetilde{u}-Av)(A\widetilde{u}-Av)'] \\ \Longleftrightarrow\; & r(A\widehat{u},\, Av) \leq r(A\widetilde{u},\, Av). \end{aligned}$$

Ex2: If $\widehat{\eta}$ is the best estimator for η in a class where all estimators are unbiased, then $r(\widehat{\eta},\, \eta) \leq r(\widetilde{\eta},\, \eta)$ for all $\widetilde{\eta}$ in the class, which is equivalent to $\mathrm{Cov}(\widehat{\eta}) \leq \mathrm{Cov}(\widetilde{\eta})$ for all $\widetilde{\eta}$ in the class. So $\widehat{\eta}$ is the uniformly (over all parameter points) minimum variance-covariance matrix estimator in the class.

Ex3: If $\widehat{Y}_f$ is the best predictor for Y_f in a class where all predictors are unbiased, then $r(\widehat{Y}_f,\, Y_f) \leq r(\widetilde{Y}_f,\, Y_f)$ for all $\widetilde{Y}_f$ in the class, which is equivalent to $\mathrm{Cov}(\widehat{Y}_f - Y_f) \leq \mathrm{Cov}(\widetilde{Y}_f) - Y_f)$.