L01 Generalized least square estimators

1. Metric system in R^n

   (1) $\langle x, y\rangle_U$
   With positive definite matrix $U \in R^{n\times n}$, for $x, y \in R^n$, $\langle x, y\rangle_U = y'Ux$ is an inner product since
   (i) $\langle x, x\rangle_U \geq 0$ and $\langle x, x\rangle_U = 0 \Longleftrightarrow x = 0$
   (ii) $\langle x, y\rangle_U = \langle y, x\rangle_U$
   (iii) $\langle \alpha x + \beta y, z\rangle_U = \alpha\langle x, z\rangle_U + \beta\langle y, z\rangle_U$.
   The norm induced is $\|x\|_U = \sqrt{\langle x, x\rangle_U} = \sqrt{x'Ux}$.

   (2) A special case
   When $U = I_n$, $\langle x, y\rangle_I = \langle x, y\rangle = y'x$ is the Frobenius inner product with norm $\|x\|_I = \sqrt{x'x} = \|\cdot\|$.

   (3) Relation
   $\langle x, y\rangle_U = y'Ux = \left(U^{1/2}y\right)'\left(U^{1/2}x\right) = \left\langle U^{1/2}x, U^{1/2}y\right\rangle$.
   $\|x\|_U^2 = x'Ux = \left(U^{1/2}x\right)'\left(U^{1/2}x\right) = \left\|U^{1/2}x\right\|^2$

2. Generalized least square estimators

   (1) Model
   Consider model $Y = X\beta + \epsilon$ with $E(\epsilon) = 0 \in R^n$.
   Models with $\epsilon \sim (0, \sigma^2 V)$ or $\epsilon \sim N(0, \sigma^2 V)$ are examples for $E(\epsilon) = 0$.

   (2) Generalized least square estimators
   $Y = X\beta + \epsilon \Longrightarrow \epsilon = Y - X\beta \Longrightarrow \|\epsilon\|_U^2 = \|Y - X\beta\|_U^2$.
   If $\|Y - X\beta\|_U^2$ is minimized by $\widehat{\beta}$, then $\widehat{\beta}$ is called a generalized least square estimator for β with respect to the norm $\|\cdot\|_U$.
   The collection of all such estimators is denoted by $\text{GLSE}_U(\beta)$.
   Clearly, $\text{GLSE}_I(\beta) = \text{OLSE}(\beta)$.

   (3) $\text{GLSE}_U(\beta)$
   Theorem: $\text{GLSE}_U(\beta) = \left(U^{1/2}X\right)^+ U^{1/2}Y + \mathcal{N}(X)$.

   Proof.

$$\begin{aligned}
& \widehat{\beta} \in \text{GLSE}_U(\beta) \Longleftrightarrow \|Y - X\widehat{\beta}\|_U^2 \leq \|Y - X\beta\|_U^2 \text{ for all } \beta \\
\Longleftrightarrow\ & \|U^{1/2}Y - U^{1/2}X\widehat{\beta}\|^2 \leq \|U^{1/2}Y - U^{1/2}X\beta\|^2 \text{ for all } \beta. \\
\Longleftrightarrow\ & U^{1/2}X\widehat{\beta} = \pi(U^{1/2}Y|\mathcal{R}(U^{1/2}X)) = U^{1/2}X\left(U^{1/2}X\right)^+ U^{1/2}Y \\
\Longleftrightarrow\ & U^{1/2}X\left[\widehat{\beta} - \left(U^{1/2}X\right)^+ U^{1/2}Y\right] = 0 \\
\Longleftrightarrow\ & X\left[\widehat{\beta} - \left(U^{1/2}X\right)^+ U^{1/2}Y\right] = 0 \\
\Longleftrightarrow\ & \widehat{\beta} - \left(U^{1/2}X\right)^+ U^{1/2}Y \in \mathcal{N}(X) \\
\Longleftrightarrow\ & \widehat{\beta} \in \left(U^{1/2}X\right)^+ U^{1/2}Y + \mathcal{N}(X).
\end{aligned}$$

   Ex1: $\text{OLSE}(\beta) = \text{GLSE}_I(\beta) = X^+Y + \mathcal{N}(X)$.

3. Minimum norm GLSE

(1) The role of $\widehat{\beta} = \left(U^{1/2}X\right)^{+} U^{1/2}Y$ in the expression for $\mathrm{GLSE}_U(\beta)$ can be replaced by any vector in $\mathrm{GLSE}_U(\beta)$, i.e.,

$$\text{If } \widetilde{\beta} \in \mathrm{GLSE}_U(\beta), \text{ then } \mathrm{GLSE}_U(\beta) = \widetilde{\beta} + \mathcal{N}(X).$$

Proof. If $\widetilde{\beta} \in \mathrm{GLSE}_U(\beta) = \widehat{\beta} + \mathcal{N}(X)$, then $\widetilde{\beta} - \widehat{\beta} \in \mathcal{N}(X)$.

For $\mathrm{GLSE}_U(\beta) = \widetilde{\beta} + \mathcal{N}(X)$, we only show $\subset$.

If $\widehat{\eta} \in \mathrm{OLSE}_U(\beta) = \widehat{\beta} + \mathcal{N}(X)$, then $\widehat{\eta} = \widehat{\beta} + u$ where $u \in \mathcal{N}(X)$.

So $\widehat{\eta} = \widetilde{\beta} + [u - (\widetilde{\beta} - \widehat{\beta})] \in \widetilde{\beta} + \mathcal{N}(X)$ since $u - (\widetilde{\beta} - \widehat{\beta}) \in \mathcal{N}(X)$.

Hence $\mathrm{GLSE}_U(\beta) \subset \widetilde{\beta} + \mathcal{N}(X)$. Similarly one can show $\supset$.

Ex2: In $\mathrm{OLSE}(\beta) = X^{+}Y + \mathcal{N}(X)$, $X^{+}Y$ can be replaced by any $\widetilde{\beta}$ in $\mathrm{OLSE}(\beta)$.

(2) $\widehat{\beta} = \left(U^{1/2}X\right)^{+} U^{1/2}Y \perp \mathcal{N}(X)$

Proof. $\widehat{\beta} \perp \mathcal{N}(X)$ means $\widehat{\beta} \perp u$ for all $u \in \mathcal{N}(X)$.

Note that $\widehat{\beta} \in R^p$, $\mathcal{N}(X) \subset R^p$, and in R^p no special inner product was defined. Hence Frobennius inner product is implied. For $u \in \mathcal{N}(X)$,

$$\begin{aligned}\langle \widehat{\beta}, u\rangle &= u'\widehat{\beta} = u'\left(U^{1/2}X\right)^{+} U^{1/2}Y = u'X^{+}X\left(U^{1/2}X\right)^{+} U^{1/2}Y \\ &= u'X'\left(X^{+}\right)'\left(U^{1/2}X\right)^{+} U^{1/2}Y = (Xu)'\left(X^{+}\right)'\left(U^{1/2}X\right)^{+} U^{1/2}Y = 0.\end{aligned}$$

Comment: $(AB)^{+} = B^{+}B(AB)^{+}$, $(AB)^{+} = (AB)^{+}AA^{+}$, $(AB)^{+} = B^{+}B(AB)^{+}AA^{+}$.

Ex3: In $\mathrm{OLSE}(\beta) = X^{+}Y + \mathcal{N}(X)$, $X^{+}Y \perp \mathcal{N}(X)$.

(3) Minimum norm GLSE

Among all vectors in $\mathrm{GLSE}_U(\beta)$, $\widehat{\beta} = \left(U^{1/2}X\right)^{+} U^{1/2}Y$ has minimum norm, i.e.,

$$\|\widehat{\beta}\|^2 \le \|\widetilde{\beta}\|^2 \text{ for all } \widetilde{\beta} \in \mathrm{GLSE}_U(\beta).$$

Proof. If $\widetilde{\beta} \in \mathrm{GLSE}_U(\beta) = \widehat{\beta} + \mathcal{N}(X)$, then $\widetilde{\beta} = \widehat{\beta} + u$ where $u \in \mathcal{N}(X)$.

By Pythagorean theorem

$$\|\widetilde{\beta}\|^2 = \|\widehat{\beta} + u\|^2 = \|\widehat{\beta}\|^2 + \|u\|^2 \ge \|\widehat{\beta}\|^2.$$

Comments: $\mathrm{GLSE}_U(\beta) = \widehat{\beta} + \mathcal{N}(X)$ contain a unique estimator if and only if $\mathcal{N}(X) = \{0\}$. In that case the unique estimator is $\widehat{\beta}$.

Ex4: $X^{+}Y$ is the minimum norm OLSE for β. If $\mathrm{OLSE}(\beta)$ has a unique estimator, then this estimator is $X^{+}Y$.

4. For HW01

(1) Affine set

For x, y in a LS V, $\alpha x + (1-\alpha)y$ is called an affine combination of x and y.

If $\mathcal{A}$ is a set in V and it is closed under affine combinations, then it is called an affine set.

(2) Subspace

A set S in LS V is a subspace if it is closed under linear combinations.

(3) Projection

V is a space where inner product $\langle\cdot,\cdot\rangle$ is defined. For $x \in V$ and subspace $S \subset V$, the projection of x onto S, $\pi(x|S)$, is defined by

$$\widehat{x} = \pi(x|S) \overset{def}{\Longleftrightarrow} \widehat{x} \in S \text{ and } \|x - \widehat{x}\|^2 \le \|x - y\|^2 \text{ for all } y \in S$$

with sufficient and necessary conditons

$$\widehat{x} = \pi(x|S) \Longleftrightarrow \widehat{x} \in S \text{ and } \langle x - \widehat{x}, y\rangle = 0 \text{ for all } y \in S.$$