L11: A biased estimator: Stein estimator

1. A biased estimator: Stein estimator

   (1) Definition
   For $y = X\beta + e$, $e \sim N(0,\, \sigma^2 I_n)$, the BLUE for β is

$$\widehat{\beta} = (X'X)^{-1}X'y \sim N(\beta,\, \sigma^2(X'X)^{-1}).$$

   To reduce the variance-covariance matrix, with $0 < c < 1$ let

$$\widehat{\beta}(c) = c\widehat{\beta} \sim N(c\beta,\, c^2\sigma^2(X'X)^{-1}).$$

   $\widehat{\beta}(c)$ is called a Stein estimator for β.

   (2) Properties
   $\widehat{\beta}(c)$ is a linear, biased estimator for β with reduced norm and variance-covariance matrix from that of the BLUE $\widehat{\beta}$ for β

   Proof : $\widehat{\beta}(c) = c(X'X)^{-1}X'y$ is a linear function of y.
   $E(\widehat{\beta}(c)) = c\beta \neq \beta$. So $\widehat{\beta}(c)$ is biased with $\|E(\widehat{\beta}(c)) - \beta\|^2 = (c-1)^2\|\beta\|^2$.
   $\|\widehat{\beta}(c)\| = \|c\widehat{\beta}\| = c\|\widehat{\beta}\| < \|\widehat{\beta}\|$. $\mathrm{Cov}(\widehat{\beta}(c)) = c^2\mathrm{Cov}(\widehat{\beta}) \leq \mathrm{Cov}(\widehat{\beta})$.

   (3) MSE of $\widehat{\beta}(c)$
   $\mathrm{MSE}(\widehat{\beta}(c)) = \|E(\widehat{\beta}(c) - \beta\|^2 + \mathrm{tr}[\mathrm{Cov}(\widehat{\beta}(c))] = (c-1)^2\|\beta\|^2 + c^2\sigma^2\mathrm{tr}[(X'X)^{-1}]$.

   (4) Conditions for $\mathrm{MSE}(\widehat{\beta}(c)) \leq \mathrm{MSE}(\widehat{\beta})$

$$\mathrm{MSE}[\widehat{\beta}(c)] \leq \mathrm{MSE}(\widehat{\beta}) \Longleftrightarrow \mathrm{MSE}[\widehat{\beta}(c)] - \mathrm{MSE}(\widehat{\beta}) \leq 0$$

$$\Longleftrightarrow \quad [\,\|\beta\|^2 + \sigma^2\mathrm{tr}(X'X)^{-1}\,]c^2 - 2c\|\beta\|^2 + \|\beta\|^2 - \sigma^2\mathrm{tr}(X'X)^{-1} \leq 0$$

$$\Longleftrightarrow \quad c \in \frac{2\|\beta\|^2 \pm \sqrt{4\|\beta\|^4 - 4\,[\,\|\beta\|^2 + \sigma^2\mathrm{tr}(X'X)^{-1}\,]\,[\,\|\beta\|^2 - \sigma^2\mathrm{tr}(X'X)^{-1}\,]}}{2[\,\|\beta\|^2 + \sigma^2\mathrm{tr}(X'X)^{-1}\,]}$$

$$\Longleftrightarrow \quad c \in \frac{2\|\beta\|^2 \pm \sqrt{4\,[\,\sigma^2\mathrm{tr}(X'X)^{-1}\,]^2}}{2[\,\|\beta\|^2 + \sigma^2\mathrm{tr}(X'X)^{-1}\,]} \Longleftrightarrow c \in \frac{\|\beta\|^2 \pm \sigma^2\mathrm{tr}(X'X)^{-1}}{\|\beta\|^2 + \sigma^2\mathrm{tr}(X'X)^{-1}}$$

$$\Longleftrightarrow \quad \frac{\|\beta\|^2 - \sigma^2\mathrm{tr}(X'X)^{-1}}{\|\beta\|^2 + \sigma^2\mathrm{tr}(X'X)^{-1}} \leq c \leq 1$$

   But it is known that $0 < c < 1$. Thus

$$\mathrm{MSE}[\widehat{\beta}(c)] \leq \mathrm{MSE}(\widehat{\beta}) \Longleftrightarrow \begin{cases} \frac{\|\beta\|^2 - \sigma^2\mathrm{tr}(X'X)^{-1}}{\|\beta\|^2 + \sigma^2\mathrm{tr}(X'X)^{-1}} \leq c < 1 & \frac{\|\beta\|^2 - \sigma^2\mathrm{tr}(X'X)^{-1}}{\|\beta\|^2 + \sigma^2\mathrm{tr}(X'X)^{-1}} > 0 \\ 0 < c < 1 & \frac{\|\beta\|^2 - \sigma^2\mathrm{tr}(X'X)^{-1}}{\|\beta\|^2 + \sigma^2\mathrm{tr}(X'X)^{-1}} \leq 0 \end{cases}$$

   The lower bound for c depends on parameters β and σ^2.

2. Mixed estimator

   (1) Two estimators

   In model $y = X\beta + e$, $e \sim N(0,\, \sigma^2 I_n)$ the BLUE for β is

$$\widehat{\beta} = (X'X)^{-1}X'y \sim N(\beta,\, \sigma^2(X'X)^{-1}).$$

Model $y_0 = X_0\beta + e_0$, $e_0 \sim N(0, \Sigma_0)$ implies
$\Sigma_0^{-1/2} y_0 = \Sigma_0^{-1/2} X_0\beta + \Sigma_0^{-1/2} e_0$, $\Sigma_0^{-1/2} e_0 \sim N(0, I)$. So the BLUE for β is

$$\widetilde{\beta} = (X_0'\Sigma_0^{-1}X_0)^{-1}X_0'\Sigma_0^{-1}y_0 \sim N(\beta, (X_0'\Sigma_0^{-1}X_0)^{-1}).$$

(2) Two risks
$\text{MSEM}(\widehat{\beta}) = \text{Cov}(\widehat{\beta}) = \sigma^2(X'X)^{-1}$, $\quad \text{MSEM}(\widetilde{\beta}) = \text{Cov}(\widetilde{\beta}) = (X_0'\Sigma_0^{-1}X_0)^{-1}$
$\text{MSE}(\widehat{\beta}) = \text{tr}[\text{Cov}(\widehat{\beta})] = \text{tr}[\sigma^2(X'X)^{-1}]$
$\text{MSE}(\widetilde{\beta}) = \text{tr}[\text{Cov}(\widetilde{\beta})] = \text{tr}[(X_0'\Sigma_0^{-1}X_0)^{-1}]$.

(3) Mixed estimator
In Model

$$\begin{pmatrix} y \\ y_0 \end{pmatrix} = \begin{pmatrix} X \\ X_0 \end{pmatrix}\beta + \begin{pmatrix} e \\ e_0 \end{pmatrix}, \begin{pmatrix} e \\ e_0 \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \Sigma_0 \end{pmatrix}\right)$$

the BLUE for β is

$$\begin{aligned}
\widehat{\beta}_m &= \left[\begin{pmatrix} X \\ X_0 \end{pmatrix}' \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \Sigma_0 \end{pmatrix}^{-1} \begin{pmatrix} X \\ X_0 \end{pmatrix}\right]^{-1} \begin{pmatrix} X \\ X_0 \end{pmatrix}' \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \Sigma_0 \end{pmatrix}^{-1} \begin{pmatrix} y \\ y_0 \end{pmatrix} \\
&= \left[\tfrac{X'X}{\sigma^2} + X_0'\Sigma_0^{-1}X_0\right]^{-1} \left[\tfrac{X'y}{\sigma^2} + X_0'\Sigma_0^{-1}y_0\right] \\
&= \left[\tfrac{X'X}{\sigma^2} + X_0'\Sigma_0^{-1}X_0\right]^{-1} \left[\tfrac{X'X}{\sigma^2}\widehat{\beta} + X_0'\Sigma_0^{-1}X_0\widetilde{\beta}\right]
\end{aligned}$$

$\widehat{\beta}_m$ is called a mixed estimator for β.

3. initial analysis

(1) $\widehat{\beta}$ and $\widetilde{\beta}$
Let $W_1 = \frac{X'X}{\sigma^2}$ and $W_2 = X_0'\Sigma_0^{-1}X_0$.
Then BLUE $\widehat{\beta}$ has $\text{MSEM}(\widehat{\beta}) = W_1^{-1}$ and $\text{MSE}(\widehat{\beta}) = \text{tr}(W_1^{-1})$.
BLUE $\widetilde{\beta}$ has $\text{MSEM}(\widetilde{\beta}) = W_2^{-1}$ and $\text{MSE}(\widetilde{\beta}) = \text{tr}(W_2^{-1})$.

(2) Mixed estimator $\widetilde{\beta}_m$
$\widehat{\beta}_m = (W_1 + W_2)^{-1}(W_1\widehat{\beta} + W_2\widetilde{\beta})$ is a weighted average of $\widehat{\beta}$ and $\widetilde{\beta}$.
$\widehat{\beta}_m$ is a BLUE for β. Thus it is a linear unbiased estimator for β with

$$\begin{aligned}
\text{MSEM}(\widehat{\beta}_m) &= (W_1 + W_2)^{-1}W_1\left(\tfrac{X'X}{\sigma^2}\right)^{-1}W_1(W_1 + W_2)^{-1} + \\
&\quad (W_1 + W_2)^{-1}W_2(X_0'\Sigma_0^{-1}X_0)^{-1}X_2(W_1 + W_2)^{-1}) \\
&= (W_1 + W_2)^{-1}W_1(W_1 + W_2)^{-1} + (W_1 + W_2)^{-1}W_2(W_1 + W_2)^{-1} \\
&= (W_1 + W_2)^{-1}.
\end{aligned}$$

$\text{MSE}(\widehat{\beta}_m) = \text{tr}[(W_1 + W_2)^{-1}]$.

L12: Unbiased mixed estimator and Bayesian estimator

1. Unbiased mixed estimator

   (1) Mixed estimator
   $\widehat{\beta}$ is BLUE of β in $y = X\beta + e$, $e \sim N(0, \sigma^2 I)$;
   $\widetilde{\beta}$ is BLUE of β in $y_0 = X_0\beta + e_0$, $e_0 \sim N(0, \Sigma_0)$;
   Mixed estimator $\widehat{\beta}_m$ is BLUE of β in $\begin{pmatrix} y \\ y_0 \end{pmatrix} = \begin{pmatrix} X \\ X_0 \end{pmatrix}\beta + \begin{pmatrix} e \\ e_0 \end{pmatrix}$, $\begin{pmatrix} e \\ e_0 \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma^2 I & 0 \\ 0 & \Sigma_0 \end{pmatrix}\right)$.

   (2) MSEM
   Let $W_1 = \frac{X'X}{\sigma^2} > 0$ and $W_2 = X_0'\Sigma_0^{-1}X_0 > 0$.
   Then $\text{MSEM}(\widehat{\beta}) = W_1^{-1}$, $\text{MSEM}(\widetilde{\beta}) = W_2^{-1}$, $\widehat{\beta}_m = (W_1 + W_2)^{-1}(W_1\widehat{\beta} + W_2\widetilde{\beta})$ and $\text{MSEM}(\widehat{\beta}_m) = (W_1 + W_2)^{-1}$.

   (3) A tool
   $(A + C'BC)^{-1} = A^{-1} - A^{-1}C'\left(B^{-1} + CA^{-1}C'\right)^{-1} CA^{-1}$

   Proof $(A + C'BC)[A^{-1} - A^{-1}C'\left(B^{-1} + CA^{-1}C'\right)^{-1} CA^{-1} = \cdots = I$.

   Comment: With $A = \Lambda$, $B = K$ and $C = I$ we have the tool used in the discussion on Ridge estimator

   $$(\Lambda + K)^{-1} = \Lambda^{-1} - \Lambda^{-1}(\Lambda^{-1} + K^{-1})^{-1}\Lambda^{-1}.$$

   (4) Better performance

   By (3),
   $$\begin{aligned} \text{MSEM}(\widehat{\beta}_m) &= (W_1 + W_2)^{-1} = (W_1 + X_0'\Sigma_0^{-1}X_0)^{-1} \\ &= W_1^{-1} - W_1^{-1}X_0'\left(\Sigma_0 + X_0W_1^{-1}X_0'\right)^{-1} X_0W_1^{-1} \le W_1^{-1} = \text{MSEM}(\widehat{\beta}) \end{aligned}$$
   Because $A \le B \Longrightarrow \text{tr}(A) \le \text{tr}(B)$,
   $\text{MSE}(\widehat{\beta}_m) = \text{tr}[\text{MSEM}(\widehat{\beta}_m)] \le \text{tr}[\text{MSEM}(\widehat{\beta})] = \text{MSE}(\widehat{\beta})$.

2. Bayes estimator

   (1) Bayesian Statistics
   In Bayesian statistics parameter θ is treated as random and is given a prior distribution with pdf $f_\theta(\theta) \propto f_0(\theta)$.
   The distribution of data y with parameter θ is treated as the conditional distribution of y given θ with pdf $f_{y|\theta}(y)$. The likelihood function is this conditional pdf treated as a function of θ. $L(\theta) = f_{y|\theta}(y) \propto L_1(\theta)$.
   The conditional distribution of θ given y is called the posterior distribution with pdf $f_{\theta|y}(\theta)$.

   (2) Bayes estimator
   The mean of posterior distribution is a function of y. This mean is the Bayesian estimator for θ.

   (3) Finding posterior pdf
   $f_{\theta|y}(\theta) = \frac{f(\theta, y)}{f_y(y)} = \frac{f_\theta(\theta)\, f_{y|\theta}(y)}{f_y(y)} \propto f_0(\theta)L_1(\theta)$.
   Often the posterior distribution of θ can be determined from $f_{\theta|y}(\theta) \propto f_0(\theta)L_1(\theta)$.

3. Bayes estimator for β

(1) Prior distribution
Forβ in $y = X\beta + e$, $e \sim N(0,\, \sigma^2 I_n)$ assign prior $\beta \sim N(\beta_0,\, \Sigma_0)$ with

$$f_\beta(\beta) \propto \exp\left[-\frac{1}{2}(\beta-\beta_0)'\Sigma_0^{-1}(\beta-\beta_0)\right] \propto \exp\left(-\frac{1}{2}\beta'\Sigma_0^{-1}\beta + \beta'\Sigma_0^{-1}\beta_0\right) = f_0(\beta).$$

(2) Likelihood function
The likelihood function

$$\begin{aligned} L(\beta) &= f_{y|\beta}(y) \propto \exp\left[-\tfrac{1}{2}(y-X\beta)'(\sigma^2 I_n)^{-1}(y-X\beta)\right] \\ &\propto \exp\left(-\tfrac{1}{2}\beta'\tfrac{X'X}{\sigma^2}\beta + \beta'\tfrac{I}{\sigma^2}X'y\right). \end{aligned}$$

(3) Posterior pdf
$f_{\beta|y}(\beta) \propto f_0(\beta)L_1(\beta) = \exp\left[-\tfrac{1}{2}\beta'\left(\Sigma_0^{-1} + \tfrac{X'X}{\sigma^2}\right)\beta + \beta'\left(\Sigma_0^{-1}\beta_0 + \tfrac{I}{\sigma^2}X'y\right)\right].$

(4) Bayesian estimator for β
Let $W_1 = \Sigma_0^{-1}$, $W_2 = \frac{X'X}{\sigma^2}$, $\widehat{\beta} = (X'X)^{-1}X'y$ (the BLUE for β) and $\widehat{\beta}_B = (W_1+W_2)^{-1}(W_1\beta_0 + W_2\widehat{\beta})$, a weighted average of β_0 and $\widehat{\beta}$. Then

$$\begin{aligned} f_{\beta|y}(\beta) &\propto \exp\left[-\tfrac{1}{2}\beta'(W_1+W_2)\beta + \beta'(W_1\beta_0 + W_2\widehat{\beta})\right] \\ &= \exp\left[-\tfrac{1}{2}\beta'(W_1+W_2)\beta + \beta'(W_1+W_2)(W_1+W_2)^{-1}(W_1\beta_0 + W_2\widehat{\beta})\right] \\ &= \exp\left[-\tfrac{1}{2}\beta'(W_1+W_2)\beta + \beta'(W_1+W_2)\widehat{\beta}_B\right] \\ &\propto \exp\left[-\tfrac{1}{2}(\beta-\widehat{\beta}_B)'(W_1+W_2)(\beta-\widehat{\beta}_B)\right]. \end{aligned}$$

Thus $\beta|y \sim N(\widehat{\beta}_B,\, (W_1+W_2)^{-1})$ with $E(\beta|y) = \widehat{\beta}_B$.
So Bayesian estimator for β is the weighted average of β_0 and $\widehat{\beta}$ with weights $W_1 = \Sigma_0$ and $W_2 = \frac{X'X}{\sigma^2}$.