1. True model : $Y = Xb + Zr + \varepsilon$, $\mathbb{E}\varepsilon = 0$, $Var\,\varepsilon = \sigma^2 I$ underfitted model.

Mis-specified model : $Y = Xb + \varepsilon$, $\mathbb{E}\varepsilon = 0$, $Var\,\varepsilon = \sigma^2 I$

LS estimate of b is $\hat{b} = (X^TX)^{-1}X^TY$

(a) Prove $\hat{b}$ is a biased estimator of b and find the bias

$$\mathbb{E}\hat{b} = \mathbb{E}[(X^TX)^{-1}X^T(Xb + Zr + \varepsilon)]$$

$$= \mathbb{E}[(X^TX)^{-1}X^TXb + (X^TX)^{-1}X^TZr + (X^TX)^{-1}X^T\varepsilon]$$

$$= b + (X^TX)^{-1}X^TZr + 0 \neq b$$

Thus $\hat{b}$ is a biased estimator of b and the bias is $(X^TX)^{-1}X^TZr$

(b) Find the variance of $\hat{b}$

$$Var\,\hat{b} = Var[(X^TX)^{-1}X^TXb + (X^TX)^{-1}X^TZr + (X^TX)^{-1}X^T\varepsilon]$$

$$= Var[(X^TX)^{-1}X^T\varepsilon]$$

$$= (X^TX)^{-1}X^T\, Var\,\varepsilon\, [(X^TX)^{-1}X^T]^T$$

$$= \sigma^2 (X^TX)^{-1}X^T I X (X^TX)^{-1}$$

$$= \sigma^2 (X^TX)^{-1}$$

(c) Prove that S^2 is a biased estimator of σ

$$\mathbb{E}S^2 = \frac{\mathbb{E}[Y^T(I - X(X^TX)^{-1}X^T)Y]}{n - r(X)}$$

$$= \frac{\mathbb{E}[(b^TX^T + r^TZ^T + \varepsilon^T)(I - X(X^TX)^{-1}X^T)(Xb + Zr + \varepsilon)]}{n - r(X)}$$

$$= \frac{\mathbb{E}[(r^TZ^T + \varepsilon^T)(I - X(X^TX)^{-1}X^T)(Zr + \varepsilon)]}{n - r(X)}$$

$$= \frac{\mathbb{E}\varepsilon^T(I - X(X^TX)^{-1}X^T)\varepsilon}{n - r(X)} + \frac{r^TZ^T(I - X(X^TX)^{-1}X^T)Zr}{n - r(X)}$$

$$\mathbb{E}\varepsilon^T(I - X(X^TX)^{-1}X^T)\varepsilon = \mathbb{E}\,tr(\varepsilon^T(I - X(X^TX)^{-1}X^T)\varepsilon)$$
$$= tr((I - X(X^TX)^{-1}X^T)\mathbb{E}\varepsilon\varepsilon^T)$$
$$= \sigma^2 tr(I - X(X^TX)^{-1}X^T)$$
$$= [n - r(X)]\sigma^2$$

$$= \sigma^2 + \frac{r^TZ^T(I - X(X^TX)^{-1}X^T)Zr}{n - r(X)} \neq \sigma^2$$

(d) Show that S^2 overestimates σ^2

$$\mathbb{E}S^2 - \sigma^2 = \frac{r^T Z^T (I - X(X^TX)^{-1}X^T) Zr}{n - r(X)}$$

$I - X(X^TX)^{-1}X^T$ is idempotent and symmetric,
then it is semi-positive definite

then $(Zr)^T (I - X(X^TX)^{-1}X^T) Zr \geq 0$ for any Zr

$\Rightarrow \mathbb{E}S^2 - \sigma^2 \geq 0$,

Thus we can conclude that S^2 overestimates σ^2.

(e) Find $\mathbb{E}\hat{\varepsilon}$ and $Var\,\hat{\varepsilon}$

$$\hat{\varepsilon} = Y - X\hat{b} = Y - X(X^TX)^{-1}X^TY = (I - X(X^TX)^{-1}X^T)Y$$

$$\Rightarrow \mathbb{E}\hat{\varepsilon} = \mathbb{E}(I - X(X^TX)^{-1}X^T)Y$$

$$= (I - X(X^TX)^{-1}X^T)\mathbb{E}Y$$

$$= (I - X(X^TX)^{-1}X^T)(Xb + Zr)$$

$$= (I - X(X^TX)^{-1}X^T)Zr$$

$$Var\,\hat{\varepsilon} = Var\,(I - X(X^TX)^{-1}X^T)Y$$

$$= (I - X(X^TX)^{-1}X^T)(Var\,Y)(I - X(X^TX)^{-1}X^T)$$

$$= (I - X(X^TX)^{-1}X^T)\sigma^2$$

2. True model: $Y = X_1 b_1 + \varepsilon$, $\mathbb{E}\varepsilon = 0$, $Var\,\varepsilon = \sigma^2 I$ Over-fitted model.

Mis-specified model: $Y = Xb + \varepsilon$, $X = (X_1\ X_2)$ $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$

(a) What is $\mathbb{E}\hat{b}$.

$$\hat{b} = (X^TX)^{-1}X^TY \Rightarrow \mathbb{E}\hat{b} = (X^TX)^{-1}X^T\mathbb{E}Y$$

$$= (X^TX)^{-1}X^TX_1b_1$$

(b) Evaluate $\mathbb{E}S^2$

$$S^2 = \frac{Y^T(I - X(X^TX)^{-1}X^T)Y}{n - r(X)} \Rightarrow \mathbb{E}S^2 = \frac{\mathbb{E}Y^T(I-H)Y}{n-r(X)}, \quad \text{where } H = X(X^TX)^{-1}X^T$$

Since $\frac{1}{\sigma^2}(I-H)\cdot\sigma^2 I = I-H$ is idempotent and $\mathrm{rank}(I-H) = n - r(X)$

then $Y^T\left[\frac{1}{\sigma^2}(I-H)\right]Y \sim \chi^2_{(n-r,\ \lambda)}$

$$\lambda = \frac{1}{2}\mu_Y^T(I-H)\mu_Y$$

$$= \frac{1}{2}(X_1b_1)^T(I-H)(X_1b_1)$$

$$= \frac{1}{2}(Xb - X_2b_2)^T(I-H)(Xb - X_2b_2)$$

$$= \frac{1}{2}(X_2b_2)^T(I-H)(X_2b_2)$$

$$\Rightarrow \mathbb{E}S^2 = \frac{\sigma^2}{n-r(X)}\mathbb{E}Y^T\left[\frac{1}{\sigma^2}(I-H)\right]Y$$

$$= \frac{\sigma^2}{n-r(X)}\left[n-r(X) + (X_2b_2)^T(I-H)(X_2b_2)\right]$$

$$= \sigma^2\left[1 + \frac{(X_2b_2)^T(I - X(X^TX)^{-1}X)(X_2b_2)}{n-r(X)}\right]$$

(c) Evaluate $\mathrm{Var}\,\hat{b}$

$$\mathrm{Var}\,\hat{b} = \mathrm{Var}\left[(X^TX)^{-1}X^TY\right] = (X^TX)^{-1}X^T\left[\mathrm{Var}\,Y\right]X(X^TX)^{-1}$$

$$= \sigma^2(X^TX)^{-1}$$

3. Suppose $Y \sim \mathcal{N}_3(0, I)$ Find the distribution of $\frac{1}{3}\left[(Y_1-Y_2)^2 + (Y_2-Y_3)^2 + (Y_3-Y_1)^2\right]$

Denote $X = \frac{1}{3}\left[(Y_1-Y_2)^2 + (Y_2-Y_3)^2 + (Y_3-Y_1)^2\right]$

$$= \frac{1}{3}\left[2Y_1^2 + 2Y_2^2 + 2Y_3^2 - 2Y_1Y_2 - 2Y_2Y_3 - 2Y_1Y_3\right]$$

$$= \frac{1}{3}[Y_1\ Y_2\ Y_3]\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

$$= Y^T\begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}Y$$

$$= Y^T(I - \frac{1}{3}J)Y$$

$I - \frac{1}{3}J$ is idempotent (from the previous exercise) and its rank is 2.

then $X = Y^T(I - \frac{1}{3}J)Y \sim \chi^2_2$

4. full rank model $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} + \varepsilon_i$, $i = 1, \cdots, n$ $\Rightarrow Y = X\beta + \varepsilon$
$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $\sum_i x_{ij} = 0$, $\sum_i x_{ij}^2 = c$

Show that $\frac{1}{p}\sum_{j=0}^{p-1} Var\,\hat{\beta}_j$ is minimized when the columns of X are mutually orthogonal.

In matrix form, $\varepsilon \sim N_n(0, \sigma^2 I)$, $X \triangleq [1\ \vec{x}_1\ \cdots\ \vec{x}_{p-1}] =: [1\ \tilde{X}]$

then then LS estimate of $\hat{\beta}$ is $(X^TX)^{-1}X^TY$

$\Rightarrow Var\,\hat{\beta} = \sigma^2 (X^TX)^{-1}$

then $\frac{1}{p}\sum_{i=1}^{p} Var\,\hat{\beta}_i = \frac{1}{p} tr(Var\,\hat{\beta}) = \frac{\sigma^2}{p} tr((X^TX)^{-1})$

Decompose $(X^TX)^{-1}$ by eigenvalue decomposition

$(X^TX)^{-1} = Q\Lambda^{-1}Q^T$ where Q is orthonormal, i.e. $QQ^T = I$ and $Q^TQ = I$
$\Lambda = diag(\lambda_0, \cdots, \lambda_{p-1})$, λ_i is eigenvalue of X^TX

then $tr[(X^TX)^{-1}] = tr(Q\Lambda^{-1}Q^T) = tr(\Lambda^{-1})$

$= \sum_{i=0}^{p-1} \frac{1}{\lambda_i}$

$\geq p^2 (\sum_{i=0}^{p-1} \lambda_i)^{-1} = p^2 (tr(\Lambda))^{-1}$

$= p^2 (tr(X^TX))^{-1}$ (by $\sum x_{ij}^2 = c$)

$= p^2 \cdot (pc)^{-1}$

$= p/c$ is fixed and

the equality holds iff $\lambda_0 = \cdots = \lambda_{p-1} = \lambda$

then $X^TX = Q\Lambda Q^T = Q(\lambda I)Q^T = \lambda I$,

$\frac{1}{p}\sum_{j=0}^{p-1} Var(\hat{\beta}_j)$ reaches the minimum when columns of X are orthogonal.

5. Regression model $Y = Xb + \varepsilon$, $\mathbb{E}\varepsilon = 0$, $\mathrm{Var}\,\varepsilon = \sigma^2 I$, Let $\hat{b}_K = (X^TX + kI)^{-1}X^TY$

(a) Prove $\mathbb{E}\hat{b}_K = b - k(X^TX + kI)^{-1}b$

$$\mathbb{E}\hat{b}_K = \mathbb{E}(X^TX+kI)^{-1}X^TY = \mathbb{E}(X^TX+kI)^{-1}X^T(Xb+\varepsilon)$$

$$= \mathbb{E}(X^TX+kI)^{-1}[(X^TX+kI)b - kb + X^T\varepsilon]$$

$$= b - k(X^TX+kI)^{-1}b + 0$$

(b) Prove $\mathrm{Var}\,\hat{b}_K = \sigma^2 X^TX(X^TX+kI)^{-2}$

$$\mathrm{Var}\,\hat{b}_K = \mathrm{Var}[(X^TX+kI)^{-1}X^T\varepsilon]$$

$$= \sigma^2(X^TX+kI)^{-1}X^TX(X^TX+kI)^{-1}$$

$$= \sigma^2(X^TX+kI)^{-1}(X^TX+kI-kI)(X^TX+kI)^{-1}$$

$$= \sigma^2[(X^TX+kI)^{-1} - k(X^TX+kI)^{-2}]$$

$$= \sigma^2[(X^TX+kI)(X^TX+kI)^{-2} - k(X^TX+kI)^{-2}]$$

$$= \sigma^2 X^TX(X^TX+kI)^{-2}$$

(c) Show that for a fixed k, the estimator $\hat{b}_K$ is unbiased for b iff $k=0$.

$$\mathbb{E}\hat{b}_K = b - k(X^TX+kI)^{-1}b$$

$$= (I - k(X^TX+kI)^{-1})b$$

"$\Rightarrow$" if $k=0$, then $\mathbb{E}\hat{b}_K = b$, thus $\mathbb{E}\hat{b}_K$ is unbiased.

"$\Leftarrow$" if $\mathbb{E}\hat{b}_K = b$, i.e. $(I - k(X^TX+kI)^{-1})b = b$

then $k(X^TX+kI)^{-1}b = 0$, denote $A = (X^TX+kI)^{-1}$

if $k \neq 0$, then $Ab = 0$ which means b is in the null space of A

but A is full rank, then $b \equiv 0$.

thus $k=0$.

6. $Y = \theta + \varepsilon$, $\varepsilon_4 \sim \mathcal{N}(0, \sigma^2 I)$, $\sum_{i=1}^{4} \theta_i = 2\pi \Rightarrow \theta^T 1 = 2\pi$

(a) Evaluate the least squares estimates of θ_i

Consider the squared loss $L(\theta) = (Y-\theta)^T(Y-\theta)$ s.t. $1^T\theta = 2\pi$

then the Lagrange of the objective is

$$\mathcal{L}(\theta) = (Y-\theta)^T(Y-\theta) - \lambda(1^T\theta - 2\pi)$$

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \theta} = -(Y-\theta) - \lambda 1 = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda} = 1^T\theta - 2\pi = 0 \end{cases}$$

$$\Rightarrow \hat{\theta} = Y + \tfrac{1}{4}(2\pi 1 - JY) \quad \text{i.e. } \hat{\theta}_i = Y_i + \tfrac{1}{4}(2\pi - \sum_{j=1}^{4} Y_j)$$

(b) Evaluate the unbiased estimator of σ^2

Since $\varepsilon = Y - \theta \sim \mathcal{N}(0, \sigma^2 I)$

then $1^T\varepsilon = 1^T(Y-\theta) = 1^TY - 2\pi \sim \mathcal{N}(0, 4\sigma^2 I)$

Notice $4\sigma^2 = \mathrm{Var}(1^T\varepsilon) = \mathbb{E}(1^TY - 2\pi)^2$

then $\frac{1}{4}(1^TY - 2\pi)^2$, i.e. $\frac{1}{4}(\sum_{i=1}^{4} Y_i - 2\pi)^2$ is a unbiased estimator of σ^2.

(c) Derive a test statistic for the hypothesis that $H_0: \theta_1 = \theta_3$ and $\theta_2 = \theta_4$.

Define a matrix $K^T = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$

$H_0: \theta_1 = \theta_3$ and $\theta_2 = \theta_4 \Leftrightarrow H_0: K^T\theta = 0$

$$\hat{\theta} = Y + \tfrac{1}{4}(2\pi 1 - JY) \doteq (I - \tfrac{1}{4}J)Y + \tfrac{\pi}{2}1$$

$$\Rightarrow \mathbb{E}\hat{\theta} = \mathbb{E}(I - \tfrac{1}{4}J)Y + \tfrac{\pi}{2}1 = (I - \tfrac{1}{4}J)\theta + \tfrac{\pi}{2}1 = \theta$$

$$\mathrm{Var}\,\hat{\theta} = \mathrm{Var}\left[(I - \tfrac{1}{4}J)Y + \tfrac{\pi}{2}1\right] = \sigma^2(I - \tfrac{1}{4}J)$$

then $\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2(I - \frac{1}{4}J))$

Since $K^T\theta = 0$ and $K^T\sigma^2(I-\frac{1}{4}J)K = 2\sigma^2 I_2$

then $K^T\hat{\theta} \sim N(0, 2\sigma^2 I)$

thus $R := \dfrac{(K^T\hat{\theta})^T K^T\hat{\theta}}{2\sigma^2} \sim \chi^2_2$ under H_0

$$Y - \hat{\theta} = \frac{1}{4}(JY - 2\pi\mathbb{1})$$

$$\Rightarrow \mathbb{E}(Y-\hat{\theta}) = \mathbb{E}(\tfrac{1}{4}(JY - 2\pi\mathbb{1})) = \tfrac{1}{4}J\theta - 2\pi\mathbb{1} = 0$$

$$Var(Y-\hat{\theta}) = \tfrac{1}{4}J \, Var Y \, \tfrac{1}{4}J = \tfrac{\sigma^2}{4}J$$

then $Y - \hat{\theta} \sim N(0, \frac{\sigma^2}{4}J)$. $\frac{1}{4}J$ is idempotent with rank 1.

thus $E := \dfrac{(Y-\hat{\theta})^T(Y-\hat{\theta})}{\sigma^2} \sim \chi^2_1$

Finally, $F(Y) = \dfrac{R/2}{E/1} \sim F_{(2,1)}$, i.e.

$$\frac{(K^T\hat{\theta})^T K^T\hat{\theta}}{4(Y-\hat{\theta})^T(Y-\hat{\theta})} \sim F_{(2,1)} \text{ under } H_0.$$

7. The linear model can written as matrix form

Let $Y = [\underbrace{Y_1 \cdots Y_m}_{\text{type a}} \ \underbrace{Y_{m+1} \cdots Y_{2m}}_{\text{type b}} \ \underbrace{Y_{2m+1} \cdots Y_{2m+n}}_{\text{type c}}]^T = [\underset{\in\mathbb{R}^m}{\vec{Y}_a^T} \ \underset{\in\mathbb{R}^m}{\vec{Y}_b^T} \ \underset{\in\mathbb{R}^n}{\vec{Y}_c^T}]^T$

$$X = \begin{bmatrix} \mathbb{1}_m & 0 \\ \mathbb{1}_m & \mathbb{1}_m \\ \mathbb{1}_n & -2\mathbb{1}_n \end{bmatrix} \qquad \beta = \begin{bmatrix} \theta \\ \phi \end{bmatrix}$$

$$Y = X\beta + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I)$$

then $\hat{\beta} = \begin{bmatrix} \hat{\theta} \\ \hat{\phi} \end{bmatrix} = (X^TX)^{-1}X^TY$ where

$$(X^TX) = \begin{bmatrix} 2m+n & m-2n \\ m-2n & m+4n \end{bmatrix} \Rightarrow (X^TX)^{-1} = \frac{1}{m^2+13mn}\begin{bmatrix} m+4n & -m+2n \\ -m+2n & 2m+n \end{bmatrix}$$

$$X^TY = \begin{bmatrix} \mathbb{1}_m^T & \mathbb{1}_m^T & \mathbb{1}_n^T \\ 0 & \mathbb{1}_m^T & -2\mathbb{1}_n^T \end{bmatrix}\begin{bmatrix} \vec{Y}_a \\ \vec{Y}_b \\ \vec{Y}_c \end{bmatrix} = \begin{bmatrix} \mathbb{1}_m^T\vec{Y}_a + \mathbb{1}_m^T\vec{Y}_b + \mathbb{1}_n^T\vec{Y}_c \\ \mathbb{1}_m^T Y_b - 2\mathbb{1}_n^T Y_c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{2m+n} Y_i \\ \sum_{i=1}^{m} Y_{m+i} - 2\sum_{i=1}^{n} Y_{2m+i} \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\theta} \\ \hat{\phi} \end{bmatrix} = \frac{1}{m^2+13mn} \begin{bmatrix} (m+4n)\sum_{i=1}^{2m+n} Y_i + (-m+2n)\left(\sum_{i=1}^{m} Y_{m+i} - 2\sum_{i=1}^{n} Y_{2m+i}\right) \\ (-m+2n)\sum_{i=1}^{2m+n} Y_i + (2m+n)\left(\sum_{i=1}^{m} Y_{m+i} - 2\sum_{i=1}^{n} Y_{2m+i}\right) \end{bmatrix}$$

$$\text{Cov}\,\hat{\beta} = \sigma^2 (X^TX)^{-1}$$

thus $\text{Cov}(\hat{\theta}, \hat{\phi}) = 0$ when $m = 2n$.

8. $Y = X\beta + \varepsilon$ $\quad \varepsilon \sim N(0, \sigma^2 I)$ $\quad \hat{\beta} = (X^TX)^{-1}X^TY$ $\quad \tilde{\beta} = c\hat{\beta}$ $\quad c \le 1$. $\quad MSE\,\tilde{\beta} = \mathbb{E}(\tilde{\beta}-\beta)^T(\tilde{\beta}-\beta)$

(a) Prove $MSE(\tilde{\beta}) = c^2\sigma^2 tr((X^TX)^{-1}) + (c-1)^2\beta^T\beta$

We know $\mathbb{E}\hat{\beta} = \beta$ and $Var\,\hat{\beta} = \sigma^2 (X^TX)^{-1}$

then $MSE(\tilde{\beta}) = \mathbb{E}(\tilde{\beta}-\beta)^T \cdot (\tilde{\beta}-\beta)$

$$= \mathbb{E}(\tilde{\beta} - c\beta - (1-c)\beta)^T(\tilde{\beta} - c\beta - (1-c)\beta)$$

$$= \mathbb{E}(\tilde{\beta}-c\beta)^T(\tilde{\beta}-c\beta) - 2(1-c)\beta^T\mathbb{E}(\tilde{\beta}-c\beta) + (1-c)^2\beta^T\beta$$

$$= tr(\sigma^2c^2 \cdot (X^TX)^{-1}) + (1-c)^2\beta^T\beta$$

$$= \sigma^2c^2 tr((X^TX)^{-1}) + (1-c)^2\beta^T\beta$$

(b) c^* be the value of c such that $MSE(\tilde{\beta})$ is a minimum

$$\frac{\partial\, MSE(\tilde{\beta})}{\partial c} = 2\sigma^2 c\, tr((X^TX)^{-1}) - 2(1-c)\beta^T\beta = 0$$

$$\Rightarrow \quad c^* = \frac{\beta^T\beta}{\sigma^2 tr((X^TX)^{-1}) + \beta^T\beta} \quad \cdots\cdots (\dagger)$$

(c) $p = 5$, $\sigma^2 = 1$, $\beta^T = (1, 2, 3, 4, 5)$ $\quad X^TX$'s $\Lambda = diag(1, 2, 3, 4, 5)$, evaluate c^*

Plugin these value in $(\dagger)$, we have

$$c^* = \frac{\sum_{i=1}^{5} i^2}{\left(\sum_{i=1}^{5}\frac{1}{i}\right) + \left(\sum_{i=1}^{5} i^2\right)} \approx 0.9601397$$

9. True model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i \quad i=1,\cdots,5,\ \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$

Mis-specified model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , i=1,\dots,5 \ , \quad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$

Rewritten the model in matrix form.

Denote $X = [\tilde{X} \ \check{X}]$ where $\tilde{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_5 \end{bmatrix} \in \mathbb{R}^{5\times 2}$ $\quad \check{X} = \begin{bmatrix} x_1^2 & x_1^3 \\ \vdots & \vdots \\ x_5^2 & x_5^3 \end{bmatrix} \in \mathbb{R}^{5\times 2}$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T Y \Rightarrow \mathbb{E}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\left(\tilde{X}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \check{X}\begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix}\right)$$

$$= \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\check{X}\begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix}$$

then the bias is

$$\mathbb{E}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} - \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\check{X}\begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} \overset{\text{plugin}}{=\!=} \begin{bmatrix} 2\beta_2 \\ 3.4\beta_3 \end{bmatrix}$$

10. $Y = X\beta + \varepsilon \quad \mathbb{E}\varepsilon = 0 \quad Var\,\varepsilon = \sigma^2 I \qquad \hat{\beta} = (X^TX)^{-1}X^TY$

additional information $u = H\beta + r$, $\mathbb{E}r = 0$ $\ Var\, r = W$, W is known.

(a) Find the generalized least squares estimate of β using all available information

The complementary information can be integrated into the regression like

$$\begin{bmatrix} Y \\ u \end{bmatrix} = \begin{bmatrix} X \\ H \end{bmatrix}\beta + \begin{bmatrix} \varepsilon \\ r \end{bmatrix} \quad \cdots (*)$$

Denote $\tilde{Y} = \begin{bmatrix} Y \\ u \end{bmatrix}$, $\tilde{X} = \begin{bmatrix} X \\ H \end{bmatrix}\beta$, $\quad \tilde{\varepsilon} = \begin{bmatrix} \varepsilon \\ r \end{bmatrix}$, $\quad \Sigma = Var\,\tilde{\varepsilon} = \begin{bmatrix} \sigma^2 I & 0 \\ 0 & W \end{bmatrix}$

then $(*)$ can be written as $\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$

then the generalized LS estimate of β is

$$\hat{\beta}_K = (\tilde{X}^T\Sigma^{-1}\tilde{X})^{-1}\tilde{X}^T\Sigma^{-1}\tilde{Y} = \left[(X^T\ H^T)^T\begin{bmatrix} \frac{1}{\sigma^2}I & 0 \\ 0 & W^{-1} \end{bmatrix} \cdot \begin{bmatrix} X^T \\ H \end{bmatrix}\right]^{-1}(X^T\ H^T)\begin{bmatrix} \frac{1}{\sigma^2}I & 0 \\ 0 & W^{-1} \end{bmatrix}\begin{bmatrix} Y \\ u \end{bmatrix}$$

$$= \left(\frac{1}{\sigma^2}X^TX + H^TW^{-1}H\right)^{-1}\left(\frac{1}{\sigma^2}X^TY + H^TW^{-1}u\right)$$

(b) $\hat{\beta}_a = (H^T W^{-1} H)^{-1} H^T W^{-1} u$ is the generalized LS estimate of β based only (u, H)

Show $\hat{\beta}_K = w_1 \hat{\beta} + w_2 \hat{\beta}_a$ and what are w_1 and w_1 and $|w_1 + w_2|$?

$$\hat{\beta}_K = \left(\frac{1}{\sigma^2} X^T X + H^T W^{-1} H\right)^{-1} \left(\frac{1}{\sigma^2} X^T Y + H^T W^{-1} u\right)$$

$$= \left(\frac{1}{\sigma^2} X^T X + H^T W^{-1} H\right)^{-1} \cdot \frac{1}{\sigma^2} (X^T X) \boxed{(X^T X)^{-1} X^T Y} +$$
$$\left(\frac{1}{\sigma^2} X^T X + H^T W^{-1} H\right)^{-1} \cdot (H^T W^{-1} H) \boxed{(H^T W^{-1} H)^{-1} H^T W^{-1} u}$$

$$= w_1 \hat{\beta} + w_2 \hat{\beta}_a \quad \text{where} \begin{cases} w_1 = \left(\frac{1}{\sigma^2} X^T X + H^T W^{-1} H\right)^{-1} \frac{1}{\sigma^2} (X^T X) \\ w_2 = \left(\frac{1}{\sigma^2} X^T X + H^T W^{-1} H\right)^{-1} (H^T W^{-1} H) \end{cases}$$

$$w_1 + w_2 = \left(\frac{1}{\sigma^2} X^T X + H^T W^{-1} H\right)^{-1} \left(\frac{1}{\sigma^2} (X^T X) + H^T W^{-1} H\right) = I$$

thus $|w_1 + w_2| = 1$.