- Finite Sample properties of the OLS estimator

$E(\hat{\beta}|X) = \beta$ if $E(Y|X=x) = x'\beta$ & iid sampling,
$\forall \beta \in \mathbb{R}^k$ when $\mathbb{X}$ has full rank ($\Leftrightarrow (\mathbb{X}'\mathbb{X})^{-1}$ exists)

With homoskedasticity, $V(\hat{\beta}|X) = \sigma^2(\mathbb{X}'\mathbb{X})^{-1}$

With heteroskedasticity, $V(\hat{\beta}|X) = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\Omega\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}$ where $\Omega = \mathrm{diag}(\sigma^2(x_1), \dots, \sigma^2(x_N))$

$$\hat{b}_1 = \frac{\sum_{i=1}^N \hat{v}_i y_i}{\sum_{j=1}^N \hat{v}_j^2}$$, $\hat{v}_i$ is the OLS residual of the auxiliary regression of X_{1i} on $X_{2i} \sim X_{ki}$

$y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$: u_i is orthogonal to $X_{1i}, \dots, X_{ki}$

$X_{1i} = \hat{\pi}_2 X_{2i} + \cdots + \hat{\pi}_k X_{ki} + \hat{v}_i$: $\hat{v}_i$ is orthogonal to $X_{2i}, \dots, X_{ki}$

$$= \boxed{\frac{\sum_{i=1}^N \hat{v}_i x_{1i}}{\sum_{j=1}^N \hat{v}_j^2}}\beta_1 + \frac{\sum_{i=1}^N \hat{v}_i u_i}{\sum_{j=1}^N \hat{v}_j^2} = \beta_1 + \frac{\sum_{i=1}^N \hat{v}_i u_i}{\sum_{j=1}^N \hat{v}_j^2}$$

(=1 by construction)

For unbiasedness of $\hat{\beta}_1$, it should be "0"

$$V(\hat{b}_1|\mathbb{X}) = V\left(\frac{\sum_{i=1}^N \hat{v}_i u_i}{\sum_{j=1}^N \hat{v}_j^2}\middle|\mathbb{X}\right) = \frac{\sum_{i=1}^N \hat{v}_i^2 \boxed{V(u_i|\mathbb{X})}}{\left[\sum_{j=1}^N \hat{v}_j^2\right]^2}$$

// $V(u_i|x_i)$ under iid

$$\circledast = \sigma^2 \frac{\sum_{i=1}^N \hat{v}_i^2}{\left[\sum_{j=1}^N \hat{v}_j^2\right]^2} = \frac{\sigma^2}{\sum_{j=1}^N \hat{v}_j^2}$$

Under Homoskedasticity,

If $\sum_{j=1}^N \hat{v}_j^2 \Downarrow$ → $V(\hat{b}_1|\mathbb{X}) \Uparrow$ get worse.
$\sigma^2 \Uparrow$ ↗

$$= \frac{\frac{1}{N}\sigma^2}{\boxed{\frac{\sum_{j=1}^N \hat{v}_j^2}{\sum_{i=1}^N (X_{1i} - \bar{X}_1)^2}}}$$

$1 - R_i^2$ ← ; ↑ R^2 of Auxiliary equation of X_{1i}

$$\underbrace{\frac{1}{N}\sum_{i=1}^N (X_{1i} - \bar{X}_1)^2}_{\text{Sample variance of } X_{1i}}$$

Under Heteroskedasticity,

$$V(\hat{b}_1|\mathbb{X}) = V\left(\frac{\sum_{i=1}^{N}\hat{v}_{1i}u_i}{\sum_{j=1}^{N}\hat{v}_{1j}^2}\middle|\mathbb{X}\right) = \frac{\sum_{i=1}^{N}\hat{v}_{1i}^2\, V(u_i|\mathbb{X})}{\left[\sum_{j=1}^{N}\hat{v}_{1j}^2\right]^2}$$

" $V(u_i|x_i)$ Under iid

$$= \underbrace{\boxed{\frac{\frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\cdot\sigma^2(x_i)}{\frac{1}{N}\sum_{j=1}^{N}\hat{v}_{1j}^2\ \ \frac{1}{N}\sum_{j=1}^{N}\sigma^2(x_j)}}}_{\text{Sample size did not affect here}} \cdot \underbrace{\frac{\frac{1}{N}\sum_{j=1}^{N}\sigma^2(x_j)}{\sum_{j=1}^{N}\hat{v}_{1j}^2}}_{\text{Sample size impacts on this part.}}$$

Sample covariance

Expectation on the variance of sample

$$E(\sigma^2(x_i)) = E\{E(u_i^2|x_i)|x_i\} = E(u_i^2)$$

↓

- If $x_i, \hat{v}_i$ uncorrelated, 1
- If $x_i\ \hat{v}_i$ positively correlated, bigger than 1. / negatively correlated, less than 1.

Note ⟨ Sample variance of Z_{1i}, Z_{2i} ⟩

$$\frac{1}{N}\sum_{i=1}^{N}(Z_{1i}-\bar{Z}_1)(Z_{2i}-\bar{Z}_2) = \frac{1}{N}\sum_{i=1}^{N}Z_{1i}Z_{2i} - \frac{1}{N}\sum_{i=1}^{N}Z_{1i}\cdot\frac{1}{N}\sum_{i=1}^{N}Z_{2i}$$

↓ Let $Z_{1i} := \hat{v}_{1i}^2$, $Z_{2i} := \sigma^2(x_i)$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{v}_{1i}^2 - \overline{\hat{v}_{1i}^2})(\sigma^2(x_i) - \overline{\sigma^2(x_i)}) = \frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\sigma^2(x_i) - \frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\cdot\frac{1}{N}\sum_{i=1}^{N}\sigma^2(x_i)$$

$$\frac{\frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\sigma^2(x_i)}{\frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\cdot\frac{1}{N}\sum_{i=1}^{N}\sigma^2(x_i)} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\hat{v}_{1i}^2-\overline{\hat{v}_{1i}^2})(\sigma^2(x_i)-\overline{\sigma^2(x_i)})}{\frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\cdot\frac{1}{N}\sum_{i=1}^{N}\sigma^2(x_i)} + 1$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{v}_{1i}^2-\overline{\hat{v}_{1i}^2})(\sigma^2(x_i)-\overline{\sigma^2(x_i)}) > 0 \iff \frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\sigma^2(x_i) > \frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\cdot\frac{1}{N}\sum_{i=1}^{N}\sigma^2(x_i)$$

$\iff$ Sample variance > 1

i.e., $\hat{v}_{1i}$ and u_i are positively correlated.

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{v}_{1i}^2-\overline{\hat{v}_{1i}^2})(\sigma^2(x_i)-\overline{\sigma^2(x_i)}) < 0 \iff \frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\sigma^2(x_i) < \frac{1}{N}\sum_{i=1}^{N}\hat{v}_{1i}^2\cdot\frac{1}{N}\sum_{i=1}^{N}\sigma^2(x_i)$$

$\iff$ Sample covariance < 1

i.e., $\hat{v}_{1i}$ and u_i are negatively correlated.

※ Variance do not give explanations about its distribution.
Thus, we can not get the distribution for the OLS estimate with finite sample.
Therefore, we should add an important assumption about distribution.

$$U_i | X_i \sim N(0, \sigma^2)$$

$E(u_i|X_i) = 0$ Homoskedasticity.

⇒ "Asymtotically Normal"

$\hat{\beta} = \beta + (X'X)^{-1}X'u$ (linear combination of u is also Normal)

linear combination of $u \sim N$

$\downarrow$

$$\hat{\beta}|X \sim N(\underset{k\times1}{\beta}, \underbrace{\sigma^2(X'X)^{-1}}_{k\times k})$$

Variance - covariance matrix.

k-lth element of $\sigma^2(X'X)^{-1}$ is the $cov(\hat{\beta}_k, \hat{\beta}_l | X)$

- Optimality Property of the OLS estimator

OLS is BLUE ($\underset{\sim}{\text{Best}}$, Linear, Unbiased Estimator)

the smallest variance.

; OLS has the smallest variance among all linear (in Y_i) estimators that are Unbiased.

$\Rightarrow$ Considering estimating $c'\beta$ by a linear combination of $\{Y_i\}$, $\underset{1\times N}{a'}\underset{N\times 1}{\mathbb{Y}}$

Conditional unbiasedness

$$E\{a'\mathbb{Y}\mid\mathbb{X}\} = a'E\{\mathbb{X}\beta + \mathbb{u}\mid\mathbb{X}\} = \underset{1\times K}{a'\mathbb{X}}\beta = \underset{1\times K}{c'}\beta, \quad \forall\beta\in\mathbb{R}^K$$

$\Rightarrow a'\mathbb{X} = c'$ Thus, we need to minimize variance under this restriction.

$$\min_{a} V(a'\mathbb{Y}\mid\mathbb{X}) = V(a'(\mathbb{X}\beta+\mathbb{u})\mid\mathbb{X}) = \underbrace{V(a'\mathbb{X}\beta\mid\mathbb{X})}_{=0} + V(a'\mathbb{u}\mid\mathbb{X})$$

s.t $\mathbb{X}'a = c$ i.e., $a'\mathbb{X} = c'$

$$= V(a'\mathbb{u}\mid\mathbb{X}) = a'V(\mathbb{u}\mid\mathbb{X})a \underset{\uparrow}{=} \sigma^2 a'a$$

Homoskedasticity

Therefore, $\min_{a} a'a$ s.t $\mathbb{X}'a = c$ is the solution for $\min V(a'\mathbb{Y}\mid\mathbb{X})$.

$$\mathcal{L} = \underset{\sim}{\frac{1}{2}}a'a - \underset{1\times K}{\lambda'}\underset{K\times 1}{(\mathbb{X}'a - c)}$$

attached just for convenience.

$$\frac{\partial\mathcal{L}}{\partial a} = \underset{N\times 1}{a} - \underset{N\times K}{\mathbb{X}}\underset{K\times 1}{\lambda} = 0 \quad \text{i.e.} \quad a = \mathbb{X}\lambda$$

$$\frac{\partial\mathcal{L}}{\partial\lambda} = -(\mathbb{X}'a - c) = 0 \quad \text{i.e.,} \quad \mathbb{X}'a = c$$

Then, $\mathbb{X}'\mathbb{X}\lambda = c$

$\therefore \lambda = (\mathbb{X}'\mathbb{X})^{-1}c$

By the assumption $\mathbb{X}$ has full rank,

$a = \mathbb{X}\lambda = \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}c \rightarrow$ estimator is $\boxed{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y} = \text{OLS}}$

※ Generally, there are a lot of Heteroskedasticity cases in real world. In these cases, OLS is not the best estimator. Thus, we can use Weighted Least Squares (WLS) estimator

---

- Using point estimation by OLS, we can get $\hat{\beta} = \beta + (X'X)^{-1}X'u$.

Then, $\hat{\beta}|X \sim N(\beta, \sigma^2(X'X)^{-1})$

Since a linear combination of the jointly normal random variables has the Normal distribution, we have

$$\hat{b}_1|X \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{N}\hat{v}_i^2}\right) \quad \text{i.e.,} \quad \frac{\hat{b}_1 - \beta_1}{\sqrt{Var(\hat{b}_1|X)}} \sim N(0,1)$$

"unknown"

$$\text{where } V(\hat{b}_1|X) = \frac{\sigma^2}{\sum_{i=1}^{N}\hat{v}_i^2}$$

known

- To compute the confidence interval for β_1,

We should use $\hat{\sigma}^2 = \frac{1}{N-K}\sum_{i=1}^{N}\hat{u}_i^2$ where $\hat{u}_i = y_i - x_i'\hat{\beta}$

$$\frac{\hat{b}_1 - \beta_1}{\hat{\sigma}\sqrt{\frac{1}{\sum_{i=1}^{N}\hat{v}_i^2}}} = \frac{\frac{\hat{b}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}}}{\frac{\hat{\sigma}\sqrt{\frac{1}{\sum_{i=1}^{N}\hat{v}_i^2}}}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}}} = \frac{\frac{\sigma(\hat{b}_1 - \beta_1)}{\hat{\sigma}\sqrt{\frac{1}{\sum_{i=1}^{N}\hat{v}_i^2}}}}{\frac{\sigma}{\frac{\hat{\sigma}}{\sigma}}} = \frac{\frac{\hat{b}_1 - \beta_1}{\sqrt{\frac{\sigma}{\sum_{i=1}^{N}\hat{v}_i^2}}}}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}} = \frac{N(0,1)}{\sqrt{\chi^2_{(N-K)}}}$$

Independent

$\sim t(N-K)$

(Divide both numerator and denominator by $\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} = \frac{\hat{\sigma}}{\sigma}$)

- Why $N(0,1)$ and $\chi^2_{(N-K)}$ are independent?

proof)

$\hat{b} = \beta + (X'X)^{-1}X'u$

OLS residual : $\hat{u}_i = y_i - x_i'\hat{b} \Rightarrow \hat{u} = Y - X\hat{b}$

$$\hat{u} = Y - X\hat{b} = Y - X(X'X)^{-1}X'Y$$

$$= X\beta + u - X(X'X)^{-1}X'(X\beta + u)$$

$$= X\beta + u - X\beta - X(X'X)^{-1}X'u = u - X[X'X]^{-1}X'u$$

$$= \underbrace{[I_N - X(X'X)^{-1}X']}_{:= M_X} u$$

$$X(X'X)^{-1}X' := P_X$$

Note ⟨Idempotent Matrix⟩

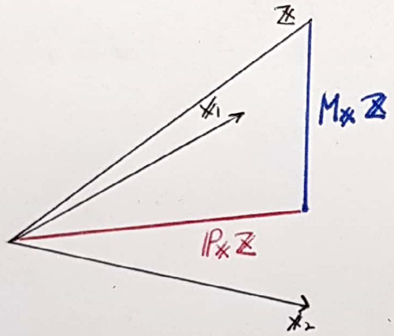
↳ $AA = A$

$M_X = I - P_X = I - X(X'X)^{-1}X'$

$$M_X \cdot M_X = (I - X(X'X)^{-1}X')(I - X(X'X)^{-1}X') = I - X(X'X)^{-1}X' = M_X$$

$P_X = X(X'X)^{-1}X'$

$$P_X \cdot P_X = X(X'X)^{-1}X' \cdot X(X'X)^{-1}X' = X(X'X)^{-1}X' = P_X$$



$\hat{b} = \beta + \underline{(X'X)^{-1}}X'u$

↳ constant. so that $X'u$ is the linear combination of u.

→ u is in the space of $M_X Z$

Under Homoskedasticity,

roughly, $E([I_N - X(X'X)^{-1}X']uu'X \mid X) = [I_N - X(X'X)^{-1}X'] \cdot \underbrace{E(uu' \mid X)}_{= \sigma^2 I_N}$

$$= \sigma^2 (X - X) = 0$$

Thus, $\underset{k\times 1}{X'u}$ and $\underset{N\times 1}{[I_N - X(X'X)^{-1}X']u}$ are orthogonal. i.e they are independent. //

(Not exactly since they have different dimensions, but "roughly orthogonal")