- Finite sample distribution of the OLS estimator (Continued)

$$\frac{\hat{\beta}_1-\beta}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum_{j=1}^{N}\hat{v}_j^2}}} \sim t(N-K)$$

$$= \frac{\dfrac{\hat{\beta}_1-\beta}{\sqrt{\sigma^2/\sum_{j=1}^{N}\hat{v}_j^2}} \sim N(0,1)}{\sqrt{\dfrac{\hat{\sigma}^2}{\sigma^2}} \sim \sqrt{\dfrac{\chi^2(N-K)}{N-K}}} \Big\} \text{ independent}$$

Let's think of it more by matrices.

$\hat{\beta}-\beta = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'u$ $\qquad \underset{N\times K}{\mathbb{X}} = (\underset{N\times 1}{\mathbb{X}_1} : \mathbb{X}_2 : \cdots : \underset{N\times 1}{\mathbb{X}_K})$

$$\mathbb{X}' = \begin{pmatrix} \mathbb{X}_1' \\ \vdots \\ \mathbb{X}_K' \end{pmatrix} \quad \text{Thus, } \mathbb{X}'u = \begin{pmatrix} \mathbb{X}_1'u \\ \vdots \\ \mathbb{X}_K'u \end{pmatrix}$$

This is from $\hat{\beta}$

Note 1 Let V = a standard normally random vector.

Then, $V'AV \sim \chi^2(\text{rank}A)$

A: Idempotent matrix

(V is a standard normal random vector so it is mutually independent)

Note 2 Idempotent matrix : $AA = A$

※ Idempotent matrix is possible to use "Spectral decomposition"

$A = H\Lambda H'$, $HH' = I$ (H: full rank)

Λ : diagonal matrix with eigen values as the diagonal elements.

Thus, $V'AV = V'(H\Lambda H')V$ $\qquad H'V \sim N(0, I_N)$

"Variance of Standard (normal) Distribution"

$E(H'VV'H) = H'E(VV')H = H'H = I$

Note1 + Note2

$V'AV = V'H\Lambda H'V$

By the fact that "The diagonal elements of Λ is only 0 or 1"

$$H'V = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$$

Therefore, $V'AV = \sum_i w_i^2$ corresponding to eigen value 1

(The # of eigen value 1 = rank A)

$\therefore$ $V'AV \sim \chi^2(\text{rank}A)$

※ Spectral Decomposition : If A is symmetric, and A has n eigenvalues,

Then $\exists\ \underbrace{[g_1, \ldots, g_n]}_{n\times n \text{ square matrix}}$ and $D = \begin{pmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_n \end{pmatrix}$

Such that $Q^{-1}AQ = D$ where $Q = [g_1 \ldots g_n]$.

To figure out that $\dfrac{\dfrac{\hat{\beta}_1-\beta_1}{\sqrt{\sigma^2/\sum_{j=1}^{N}\hat{v}_{ij}^2}}}{\sqrt{\dfrac{\hat{\sigma}^2}{\sigma^2}}}$ $\left.\begin{matrix}\sim N(0,1) \\ \sim \chi^2(N-K)\end{matrix}\right]$ are independent,

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

Thus, using matrix forms,

$$(N-K)\hat{\sigma}^2 = \underbrace{u}_{N(0,\sigma^2)}{}'(I_N - X(X'X)^{-1}X')u$$

Divide by σ^2

$$\frac{u}{\sigma} = \begin{pmatrix} \frac{u_1}{\sigma} \\ \vdots \\ \frac{u_N}{\sigma} \end{pmatrix} \sim N(0,1)$$

Each term has standard Normal dist.

$$(N-K)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{u'}{\sigma}(I_N - X(X'X)^{-1}X')\frac{u}{\sigma}$$

$$(N-K)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{u'}{\sigma}\underline{(I_N - X(X'X)^{-1}X')}\frac{u}{\sigma}$$

Let's think about this rank.

$\underline{\text{rank}(I_N - X(X'X)^{-1}X')}$

$= \text{trace}(I - X(X'X)^{-1}X') = \text{trace}(I) - \text{trace}(X(X'X)^{-1}X')$

$= N - \text{trace}((X'X)^{-1}X'X)$ (By the property of trace)

$= N - \text{trace}(I_K) = \underline{N-K}$

$$(N-K)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{u'}{\sigma}(I_N - X(X'X)^{-1}X')\frac{u}{\sigma} \sim \chi^2(N-K)$$

i.e., $\boxed{\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2(N-K)}{N-K}}$

Now, $\underline{\text{show } X'u \text{ and } I_N - X(X'X)^{-1}X' \text{ are independent.}}$

$$u'(I_N - X(X'X)^{-1}X')u = u'M_X u = u'M_X M_X u$$

$$\underset{N\times 1}{M_X u} = \begin{pmatrix} m_1' \\ \vdots \\ m_N' \end{pmatrix} u = \begin{pmatrix} m_1'u \\ \vdots \\ m_N'u \end{pmatrix}$$

By $\text{rank}(M_X) = N-K$,

$m_1' \dots m_{N-K}'$ are independent and $m_{N-K+1}', \dots, m_N'$ are linear combinations of $m_1', \dots, m_{N-K}'$

$$= \begin{pmatrix} \boxed{\begin{matrix} m_1'u \\ \vdots \\ m_{N-K}'u \end{matrix}} \\ m_{N-K+1}'u \\ \vdots \\ m_N'u \end{pmatrix}$$

Then, we show $\begin{pmatrix} m_1' u \\ \vdots \\ m_{N-K}' u \end{pmatrix}$ and $\mathbb{X}'u$ are independent.

$\Downarrow$ Stack them!

$$\begin{pmatrix} m_1' u \\ \vdots \\ m_{N-K}' u \\ \hline \mathbb{X}_1' u \\ \vdots \\ \mathbb{X}_K' u \end{pmatrix} = \begin{pmatrix} m_1' \\ \vdots \\ m_{N-K}' \\ \mathbb{X}_1' \\ \vdots \\ \mathbb{X}_K' \end{pmatrix} \underset{N\times 1}{u}$$

(*)

$M_{\mathbb{X}} \mathbb{X} = 0$ so, they are orthogonal.

$m_1, \dots, m_{N-K}$: row vectors of $M_{\mathbb{X}}$.

$\mathbb{X}_1, \dots, \mathbb{X}_K$: column vectors of $\mathbb{X}$

Thus, $m_1' \cdots m_{N-K}'$ and $\mathbb{X}_1' \cdots \mathbb{X}_K'$ are linearly independent.

$\boxed{\text{Note} \quad \text{Under Normality, } cov = 0 \Leftrightarrow \text{independent}}$

Under Homoskedasticity, check covariance b/w $\begin{pmatrix} m_1' \\ \vdots \\ m_{N-K}' \end{pmatrix}$ and $\begin{pmatrix} \mathbb{X}_1' \\ \vdots \\ \mathbb{X}_K' \end{pmatrix}$

$$E(m_j' u \, \mathbb{X}_k' u \mid \mathbb{X}) = E(m_j' u u' \mathbb{X}_k \mid \mathbb{X}) = m_j' E(u u' \mid \mathbb{X}) \mathbb{X}_k$$

$$= \sigma^2 m_j' \mathbb{X}_k = 0 \quad (\because m_j, \mathbb{X}_k \text{ are orthogonal by (*)})$$

Therefore, $\mathbb{X}'u$ and $[I_N - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}']u$ are independent. //

point

$$\frac{\dfrac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / \sum_{j=1}^{N} \hat{v}_{1j}^2}}}{\sqrt{\dfrac{\hat{\sigma}^2}{\sigma^2}}} \qquad \begin{matrix} \sim N(0,1) \\[2em] \sim \sqrt{\dfrac{\chi^2(N-K)}{N-K}} \end{matrix}$$

$\hat{\beta}_1 - \beta_1 = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'u$

$cov = 0$ under Homoske. i.e., independent.

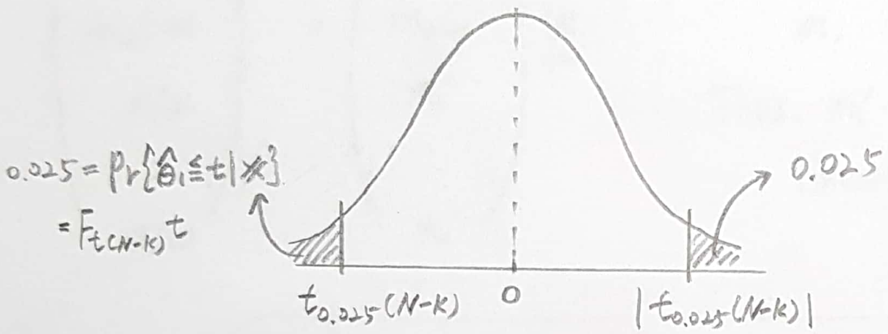
$(N-K)\hat{\sigma}^2 = u'[I_N - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}']u$

- confidence interval

$$\hat{\theta}_1 \equiv \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum_{j=1}^{N} \hat{v}_j^2}}} \sim t(N-K)$$

Note: $\Pr\{\hat{\theta}_1 \leq t \mid \mathbf{X}\} = F_{t(N-K)}(t)$

Conisider 95% confidence interval



$$t_{0.025}(N-K) \leq \frac{\hat{\beta} - \beta_1}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum_{j=1}^{N} \hat{v}_j^2}}} \leq |t_{0.025}(N-K)|$$

$$\Downarrow$$

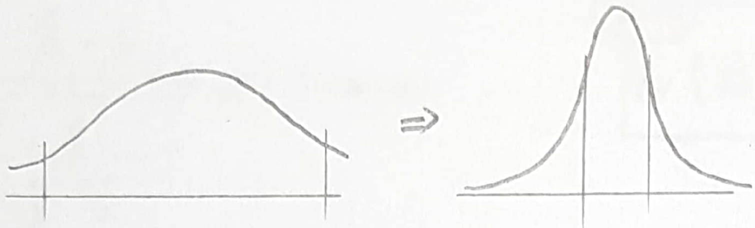
$$\boxed{\hat{\beta}_1 - |t_{0.025}(N-K)| \sqrt{\frac{\hat{\sigma}^2}{\sum_{j=1}^{N} \hat{v}_j^2}} \leq \beta_1 \leq \hat{\beta}_1 + |t_{0.025}(N-K)| \sqrt{\frac{\hat{\sigma}^2}{\sum_{j=1}^{N} \hat{v}_j^2}}}$$

→ This means, with fixed $\mathbf{X}$ (conditional on $\mathbf{X}$), u_i is not constant. Holding $\mathbf{X}$, values of u_is are different from each other by sampling.

(That is, u_i is not fixed so that it be decided by sampling)

→ There are several ways to get confidence intervals.

→ Basically all of us want to reduce the length of the interval.



- Hypothesis test

$H_0 : \beta_1 = 0$ v. $H_1 : \beta_1 \neq 0$

$$\frac{\hat{\beta}_1 - 0}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum_{j=1}^{N} \hat{v}_{ij}^2}}} \sim t(N-K)$$

→ check the lecture note.

The reason we chose here is to decide the "Power"

Note <Linear combination of coefficients>

Let $\underset{K\times 1}{C} := \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{pmatrix}$ and $\sum_{k=1}^{K} c_k \beta_k = C'\beta$ where $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}$

Then $c'\hat{\beta} = \sum_{k=1}^{K} c_k \hat{\beta}_k \sim N\left(\sum_{k=1}^{K} c_k \beta_k,\ \underbrace{\underset{1\times K}{c'}[\sigma^2 \underset{K\times N\ N\times K}{(X'X)^{-1}}]\underset{K\times 1}{c}}_{1\times 1\ (\text{scalar})}\right)$

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

From above, $c'\hat{\beta} \sim N(c'\beta, c'(\sigma^2(X'X)^{-1})c)$

$$\frac{c'\hat{\beta} - c'\beta}{\sqrt{c'[\sigma^2(X'X)^{-1}]c}} \sim N(0,1) \longrightarrow \frac{\dfrac{c'\hat{\beta} - c'\beta}{\sqrt{c'[\cancel{\sigma^2}(X'X)^{-1}]c}}}{\sqrt{\dfrac{\hat{\sigma}^2}{\cancel{\sigma^2}}}} \begin{matrix} \sim N(0,1) \\ \sim \sqrt{\dfrac{\chi^2(N-K)}{N-K}} \end{matrix} \sim t(N-K)$$

Divide by $\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}$

⇒ Therefore, using vector c, we can also easily check the unbiasedness of coefficients using t-distribution.

- Inferences about multiple linear combinations of elements of β

$$\log Wage_i = \beta_1 + \beta_2 \cdot edu_i + \beta_3 \cdot edu_i^2 + \beta_4 \cdot experience + \beta_5 \cdot experience^2 + \beta_6\, experience \cdot edu_i + \cdots$$

→ If we want to know "experience does not affect Wage",

β_4, β_5 and β_6 should be "0" (zero).

Thus, $H_0: \beta_4 = 0,\ \beta_5 = 0$ and $\beta_6 = 0$

$$\begin{pmatrix} c_1'\beta \\ \vdots \\ c_r'\beta \end{pmatrix} = \underset{r\times k}{\begin{pmatrix} c_1' \\ \vdots \\ c_r' \end{pmatrix}} \beta = \underset{r\times k\ k\times 1}{\mathbb{C}\beta} \quad \text{where } \underset{r\times k}{\mathbb{C}} := \begin{pmatrix} c_1' \\ \vdots \\ c_r' \end{pmatrix}$$

$\mathbb{C}\beta$ can be estimated by $\mathbb{C}\hat{\beta}$.

Under Normality, $\underbrace{\underset{r\times k\ k\times 1}{\mathbb{C}\hat{\beta}}}_{r\times 1} | \mathbb{X} \sim N(\mathbb{C}\beta, \underbrace{\underset{r\times k}{\mathbb{C}} \underset{k\times k}{[\sigma^2(\mathbb{X}'\mathbb{X})^{-1}]} \underset{k\times r}{\mathbb{C}'}}_{r\times r \text{ matrix.}})$

$$\mathbb{C}\hat{\beta} - \mathbb{C}\beta \,|\, \mathbb{X} \sim N(0, \mathbb{C}[\sigma^2(\mathbb{X}'\mathbb{X})^{-1}]\mathbb{C}')$$

At the previous page, recall the note.

$$\underset{1\times k}{c'\hat{\beta}} - \underset{1\times k}{c'\hat{\beta}} \sim N(0, \underbrace{c'[\sigma^2(\mathbb{X}'\mathbb{X})^{-1}]c}_{*})$$

↓ Divide by $*$

$$\frac{c'\hat{\beta} - c'\hat{\beta}}{c'[\sigma^2(\mathbb{X}'\mathbb{X})^{-1}]c} \sim N(0,1)$$

Note — ⟨Cholesky's decomposition⟩

$\mathbb{C}[\sigma^2(X'X)^{-1}]\mathbb{C}' = \Gamma\Gamma'$, Γ is a lower triangular matrix

$$\mathbb{C}\hat{\beta} - \mathbb{C}\beta \sim N(0, \mathbb{C}[\sigma^2(X'X)^{-1}]\mathbb{C}')$$

i.e., $\mathbb{C}\hat{\beta} - \mathbb{C}\beta \sim N(0, \Gamma\Gamma')$ by Cholesky's decomposition.

Multiplying Γ^{-1},

$$\Gamma^{-1}(\mathbb{C}\hat{\beta} - \mathbb{C}\beta) \sim N(0, \Gamma^{-1}(\Gamma\Gamma')(\Gamma^{-1})')$$

i.e., $N(0, I_r)$

Note — ⟨2x2 matrix cholesky's decomposition.⟩

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a^2 & ab \\ ab & c^2 \end{pmatrix}$$

$a = \sigma_1$, $b = \rho\sigma_2$ $c = \sqrt{1-\rho^2}$

$$\underbrace{(\mathbb{C}\hat{\beta} - \mathbb{C}\beta)'(\Gamma^{-1})'}_{N(0, I_r)}\underbrace{\Gamma^{-1}(\mathbb{C}\hat{\beta} - \mathbb{C}\beta)}_{N(0, I_r)}$$

$$\begin{pmatrix} w_1 & \cdots & w_r \end{pmatrix}\begin{pmatrix} w_1 \\ \vdots \\ w_r \end{pmatrix} = w_1^2 + \cdots + w_r^2 \sim \chi^2_{(r)}$$

Also, $(\Gamma^{-1})'\Gamma^{-1} = (\Gamma\Gamma')^{-1} = \underline{\{\mathbb{C}[\sigma^2(X'X)^{-1}]\mathbb{C}'\}^{-1}}$ (**)

Divide by $\frac{\hat{\sigma}^2}{\sigma^2}$ ⇒ Then, σ^2 will disappear and we know "$\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2(N-K)}{N-K}$" and "$\frac{\hat{\sigma}^2}{\sigma^2}$ and (**) are independent". Therefore, $\frac{\chi^2_{(r)}/r}{\chi^2(N-K)/N-K} \sim F(r, N-K)$

⇒ We can check the unbiasedness of $\hat{\beta}$ for multiple cases using F-distribution.

- Finite Sample Properties of $\hat{\sigma}^2$

$$E(\hat{\sigma}^2 \mid \mathbb{X}) = \frac{1}{N-K} E\{\mathbb{u}'(I_N - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')\mathbb{u} \mid \mathbb{X}\}$$

$$= \frac{1}{N-K} \text{trace } E\{\mathbb{u}'(I_N - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')\mathbb{u} \mid \mathbb{X}\}$$

$$= \frac{1}{N-K} E\{\text{trace}(\mathbb{u}'(I_N - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')\mathbb{u} \mid \mathbb{X})\}$$

$$= \frac{1}{N-K} E\{\text{trace }(I_N - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')\mathbb{u}\mathbb{u}' \mid \mathbb{X})\}$$

$$= \frac{1}{N-K} \text{trace } E\{\underbrace{(I_N - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')}_{\text{Deterministic matrix.}}\mathbb{u}\mathbb{u}' \mid \mathbb{X}\}$$

$$= \frac{1}{N-K} \text{trace }(I_N - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')\, E(\mathbb{u}\mathbb{u}' \mid \mathbb{X})$$

$$= \frac{\sigma^2}{N-K} \text{trace }(I_N - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}') = \frac{\sigma^2}{N-K}\cdot(N-K) = \sigma^2$$

Therefore, $\hat{\sigma}^2$ is a conditionally umbiased estimator.

< TA , 1/29 >

- Variance of OLS estimator

$$Var(\hat{\beta}_1 | \mathbb{X}) = \frac{\sigma^2}{\sum_{j=1}^{N} \hat{v}_{1j}^2} \quad \text{(Under Homoskedasticity)}$$

$$= \frac{\boxed{\frac{1}{N}\sum_{i=1}^{N} \hat{v}_{1i}^2 \cdot \sigma^2(x_i)}}{\underbrace{\boxed{\frac{1}{N}\sum_{j=1}^{N} \hat{v}_{1j}^2 \cdot \frac{1}{N}\sum_{i=1}^{N} \sigma^2(x_i)}}_{\exists \text{ correlation b/w } \hat{v}_{1j} \text{ and } \hat{\sigma}^2}} \cdot \frac{\frac{1}{N}\sum_{i=1}^{N} \sigma^2(x_i)}{\sum_{j=1}^{N} \hat{v}_{1j}^2} \quad \text{(Under Hetero-skedasticity)}$$

$$cov(\hat{v}_{1i}^2, \sigma^2(x_i)) = \frac{1}{N}\sum_{i=1}^{N} (\hat{v}_{1i}^2 - \overline{\hat{v}_1^2})(\sigma^2(x_i) - \overline{\sigma^2(x_i)})$$

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

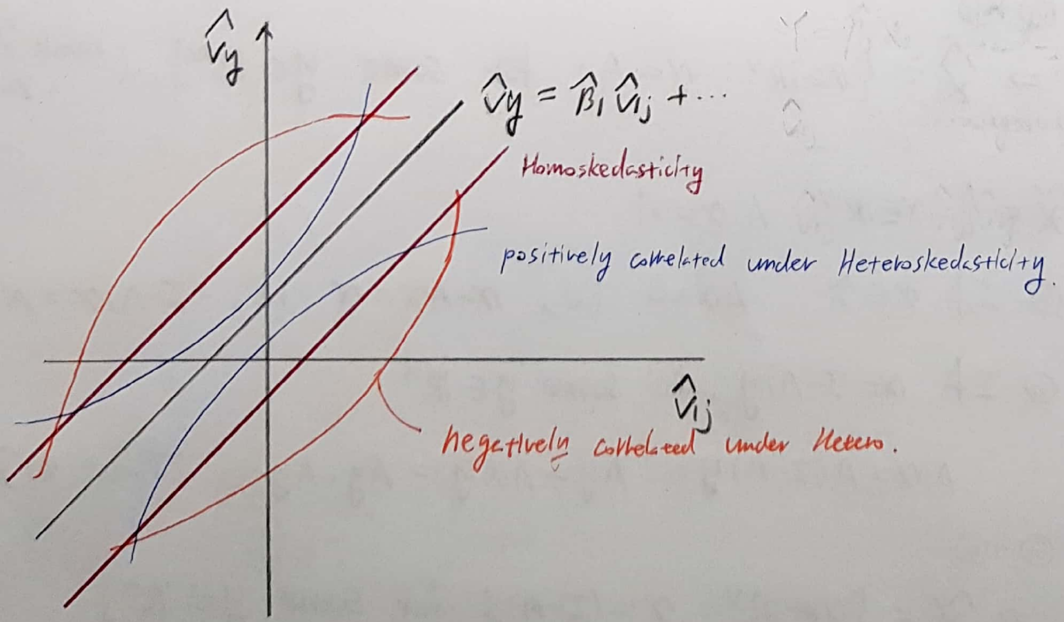
If >0, $Var(\hat{\beta}_1 | \mathbb{X})$ under Hetero is larger than Homo.

If <0, " is smaller than Homo.

Geometrically,

$$X_{1j} = \hat{\gamma}_2 X_{2j} + \cdots + \hat{v}_{1j}$$

$$Y = \hat{\pi}_2 X_{2j} + \cdots + \hat{v}_y \qquad \text{Then, } \hat{\beta}_1 = \frac{\sum_{j=1}^{N} \hat{v}_{1j}\hat{v}_y}{\sum_{j=1}^{N} \hat{v}_{1j}^2}$$



$\underset{N\times N}{A}$: idempotent matrix

① eigenvalues of A : 1 or 0

② $A = H\Lambda H'$, H : full rank

Λ : is eigenvalued vector

③ rank A = trace A

proof) ① $Ax = \lambda x$

$AAx = A\lambda x = \lambda Ax$ Then, $\left.\begin{array}{l} Ax = \lambda Ax = \lambda \cdot \lambda x \\ Ax = \lambda x \end{array}\right]$ $\lambda = \lambda^2$

Then, $\lambda = 0, 1$. //

② (i) $\hat{X} = \{x \in \mathbb{R}^N : (I-A)\cdot x = 0\}$

space of eigenvalues with eigenvalue 1

ⓐ If $x \in \hat{X}$, $(I-A)\cdot x = 0$ $x = Ax$.

ⓑ If $x = Ay$ for some $y \in \mathbb{R}^n$,

$(I-A)Ay = Ay - A\cdot Ay = Ay - Ay = 0$. Thus, $x \in \hat{X}$

ⓐ+ⓑ

$\Rightarrow \hat{X} = \{x \in \mathbb{R}^N : x = Ay$ for some $y \in \mathbb{R}^N\}$ (rank $\hat{X}$ = rank A)

(ii) $\tilde{X} = \{x \in \mathbb{R}^N : A\cdot x = 0\}$

ⓐ If $x \in \tilde{X}$, $Ax = 0$ i.e., $x - Ax = x$ i.e., $(I-A)x = x$

ⓑ If $x = (I-A)y$, for some $y \in \mathbb{R}^n$

$A\cdot x = A(I-A)y = Ay - A\cdot Ay = Ay - Ay = 0$. Thus, $x \in \tilde{X}$

ⓐ+ⓑ

$\Rightarrow \tilde{X} = \{x \in \mathbb{R}^N : x = (I-A)y$ for some $y \in \mathbb{R}^N\}$

$\hat{\mathcal{X}}$ and $\tilde{\mathcal{X}}$ are orthogonal components

$$\forall \hat{x} \in \hat{\mathcal{X}},\ \forall \tilde{x} \in \tilde{\mathcal{X}},\quad \hat{x} \circ \tilde{x} = 0$$

proof) $\hat{x} = A \cdot a$ for some $a \in \mathbb{R}^N$

$\tilde{x} = (I - A) \cdot b$ for some $b \in \mathbb{R}^N$

$$\hat{x} \circ \tilde{x} = \hat{x}'\tilde{x} = a'A'(I-A)b = a'(A - AA)b = a' \cdot 0 \cdot b = 0 \ .//$$

$$\text{Let } H := [\overbrace{H_1 \mid H_2 \mid \cdots \mid H_k}^{=: \hat{H}} \mid \overbrace{H_{k+1} \mid \cdots \mid H_N}^{=: \tilde{H}}]$$

$H_i \in \hat{\mathcal{X}}$ for $i \in \{1, \dots, k\}$ $\quad$ $H_{i'} \in \tilde{\mathcal{X}}$ for $i' \in \{k+1, \dots, N\}$

$H_i \circ H_i = 1$ $\quad$ $H_{i'} \circ H_{i'} = 1$

$H_i \circ H_j = 0$, $\forall i \neq j$ $\quad$ $H_{i'} \circ H_{j'} = 0$

$$AH = A[\hat{H}\ \ \tilde{H}] = [A\hat{H}\ \ A\tilde{H}] = [\hat{H}\ \ 0]$$

$$H'AH = \begin{bmatrix} \hat{H}' \\ \tilde{H}' \end{bmatrix} [\hat{H}\ \ 0] = \begin{pmatrix} \hat{H}'\hat{H} & \hat{H}' \cdot 0 \\ \tilde{H}'\hat{H} & \tilde{H}' \cdot 0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} = \Lambda$$

$$H'AH = \Lambda \qquad A = H'^{-1} \Lambda H^{-1}$$

$$= H \Lambda H' \quad (\text{By } (H')^{-1} = H,\ (H)^{-1} = H')$$

Then, $\operatorname{rank} A = \operatorname{rank} H \Lambda H'$

$$= \operatorname{rank} \Lambda = \operatorname{trace} \Lambda = \operatorname{trace} \Lambda \underbrace{H'H}_{=I}$$

$$= \underset{\substack{\uparrow \\ (\text{linear operator})}}{\operatorname{trace}} H \Lambda H' = \operatorname{trace} A \ .//$$