- The meaning of $\hat{\beta}_1$

$$\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y}$$

$$\hat{\mathbb{u}} = \mathbb{Y} - \mathbb{X}\hat{\beta} \quad \text{i.e., } \mathbb{Y} = \mathbb{X}\hat{\beta} + \hat{\mathbb{u}} = \hat{\beta}_1\mathbb{X}_1 + \hat{\beta}_2\mathbb{X}_2 + \cdots + \hat{\beta}_K\mathbb{X}_K + \hat{\mathbb{u}}$$

where $\underset{N\times K}{\mathbb{X}} = (\underset{N\times 1}{\mathbb{X}_1} \vdots \underset{N\times 1}{\mathbb{X}_2} \vdots \cdots \underset{N\times 1}{\mathbb{X}_K})$

$\boxed{\text{Note 1}}$ $\hat{\mathbb{u}}$ and $\mathbb{X}_1, \dots, \mathbb{X}_K$ are orthogonal $\Leftrightarrow \hat{\mathbb{u}}'\mathbb{X}_1 = \hat{\mathbb{u}}'\mathbb{X}_2 = \cdots = \hat{\mathbb{u}}'\mathbb{X}_K = 0$

To check $\hat{\beta}_1$, get <u>Auxiliary regression of $X_1$ on $X_2, \dots, X_K$</u>

→ OLS estimates of the coefficients $\hat{\alpha}_2, \dots, \hat{\alpha}_K$

$$\underset{N\times 1}{\mathbb{X}_1} = \hat{\alpha}_2\mathbb{X}_2 + \hat{\alpha}_3\mathbb{X}_3 + \cdots + \hat{\alpha}_K\mathbb{X}_K + \underset{N\times 1}{\hat{\mathbb{V}}_1}$$

$\boxed{\text{Note 2}}$ By OLS contruction, $\hat{\mathbb{V}}_1$ is orthogonal to $\mathbb{X}_2, \dots \mathbb{X}_K$
$\Leftrightarrow \hat{\mathbb{V}}_1'\mathbb{X}_2 = \hat{\mathbb{V}}_1'\mathbb{X}_3 = \cdots = \hat{\mathbb{V}}_1'\mathbb{X}_K = 0$

Rearrange!

$$\hat{\mathbb{V}}_1 = \mathbb{X}_1 - \hat{\alpha}_2\mathbb{X}_2 - \hat{\alpha}_3\mathbb{X}_3 - \cdots - \hat{\alpha}_K\mathbb{X}_K$$

↳ multiply $\hat{\mathbb{u}}'$

$$\hat{\mathbb{u}}'\hat{\mathbb{V}}_1 = \hat{\mathbb{u}}'\mathbb{X}_1 - \hat{\alpha}_2\hat{\mathbb{u}}'\mathbb{X}_2 - \hat{\alpha}_3\hat{\mathbb{u}}'\mathbb{X}_3 - \cdots - \hat{\alpha}_K\hat{\mathbb{u}}'\mathbb{X}_K = 0 \quad \text{(by note 1)}$$

So, <u>$\hat{\mathbb{u}}'\hat{\mathbb{V}}_1 = 0$</u> (*)

multiply $\hat{\mathbb{V}}_1'$

$$\hat{\mathbb{V}}_1'\mathbb{X}_1 = \hat{\alpha}_2\hat{\mathbb{V}}_1'\mathbb{X}_2 + \hat{\alpha}_3\hat{\mathbb{V}}_1'\mathbb{X}_3 + \cdots + \hat{\alpha}_K\hat{\mathbb{V}}_1'\mathbb{X}_K + \hat{\mathbb{V}}_1'\hat{\mathbb{V}}_1 = 0 + \cdots + 0 + \hat{\mathbb{V}}_1'\hat{\mathbb{V}}_1 \quad \text{(By note 2)}$$

So, <u>$\hat{\mathbb{V}}_1'\mathbb{X}_1 = \hat{\mathbb{V}}_1'\hat{\mathbb{V}}_1$</u> (**)

Then, <u>multiply $\hat{\mathbb{V}}_1'$ to the $\mathbb{Y}$.</u>

$$\hat{\mathbb{V}}_1'\mathbb{Y} = \hat{\mathbb{V}}_1(\mathbb{X}\hat{\beta} + \hat{\mathbb{u}}) = \hat{\mathbb{V}}_1'(\hat{\beta}_1\mathbb{X}_1 + \hat{\beta}_2\mathbb{X}_2 + \cdots + \hat{\beta}_K\mathbb{X}_K + \mathbb{u})$$

$$= \hat{\beta}_1\hat{\mathbb{V}}_1'\mathbb{X}_1 + \underbrace{\hat{\beta}_2\hat{\mathbb{V}}_1'\mathbb{X}_2 + \cdots + \hat{\beta}_K\hat{\mathbb{V}}_1'\mathbb{X}_K}_{=0 \text{ (By Note 2)}} + \underbrace{\hat{\mathbb{V}}_1'\mathbb{u}}_{=0\,(*)} = \hat{\beta}_1\hat{\mathbb{V}}_1'\mathbb{X}_1 \underset{\text{By }(**)}{=} \hat{\beta}_1\hat{\mathbb{V}}_1'\hat{\mathbb{V}}_1$$

Thus, $\hat{\mathbb{V}}_1'\mathbb{Y} = \hat{\beta}_1\hat{\mathbb{V}}_1'\hat{\mathbb{V}}_1$. If $\hat{\mathbb{V}}_1'\hat{\mathbb{V}}_1 \neq 0$, $\boxed{\hat{\beta}_1 = \dfrac{\hat{\mathbb{V}}_1'\mathbb{Y}}{\hat{\mathbb{V}}_1'\hat{\mathbb{V}}_1} = \dfrac{\hat{\mathbb{V}}_1'\hat{\mathbb{V}}_Y}{\hat{\mathbb{V}}_1'\hat{\mathbb{V}}_1}}$

<u>Regress $\mathbb{Y}$ on $X_2, \dots, X_K$</u> : $\mathbb{Y} = \hat{\pi}_2\mathbb{X}_2 + \hat{\pi}_3\mathbb{X}_3 + \cdots + \hat{\pi}_K\mathbb{X}_K + \hat{\mathbb{V}}_Y$, Then $\boxed{\hat{\mathbb{V}}_1'\mathbb{Y} = \hat{\mathbb{V}}_1'\hat{\mathbb{V}}_Y}$

- What OLS estimator estimates if the linear regression model does not hold?

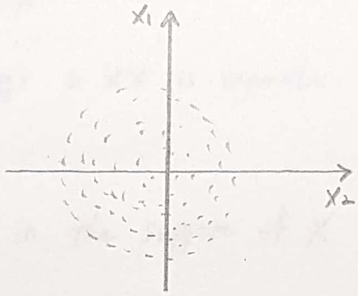
$$\min E\{[Y-g(x)]^2\} \iff \min\{[Y-E(Y|X)]^2\} + E\{[E(Y|X)-g(x)]^2\}$$

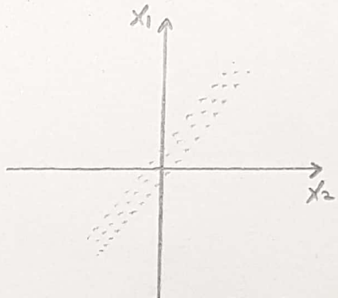
when $g(x) = X'b$

$$\iff \min_b E\{[E(Y|X)-X'b]^2\}$$

※ OLS estimator chooses the best approximation to $E(Y|X)$ within a functional form $X'b$ using the mean squared loss function.

$$E([E(Y|X)-X'b]^2) = \int_{x\in Supp(X)} [E(Y|X=x)-x'b]^2 f(x)\,dx$$




OLS well-defined

- Fit of the model measure $R^2$

$$Y = \beta_1 X_1 + \cdots + \beta_K X_K + u$$

$$Var(Y) = Var(\beta_1 X_1 + \cdots + \beta_K X_K) + Var(u) + \underline{Cov(\beta_1 X_1 + \cdots + \beta_K X_K,\ u)}$$

$$= \beta_1 \underline{Cov(X_1, u)} + \cdots + \beta_K Cov(X_K, u)$$

$$\Downarrow$$

$$\underbrace{E(X_1 u)}_{=0} - E(X_1)\cdot \underbrace{E(u)}_{=0} = 0$$

$$E(X_1 u) = E(E(X_1 u|X)) = E(X_1 \cdot E(u|X)) = 0$$

$X_1$ is constant

Thus, $Cov(\beta_1 X_1 + \cdots + \beta_K X_K,\ u) = 0$

Divide by $Var(Y)$,

$$1 = \boxed{\frac{Var(\beta_1 X_1 + \cdots + \beta_K X_K)}{Var(Y)}} + \frac{Var(u)}{Var(Y)}$$

$$\hat{y}_i = \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_K x_{Ki}$$

$$R^2 = \frac{\frac{1}{N}\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}{\frac{1}{N}\sum_{i=1}^N (y_i - \bar{y})^2}$$

where $\bar{\hat{y}} = \frac{1}{N}\sum_{i=1}^N \hat{y}_i$

$\bar{y} = \frac{1}{N}\sum_{i=1}^N y_i$

$\boxed{\text{Note}}$ If there is a constant term, in the model, then $\bar{\hat{y}} = \bar{y}$

$$\bar{\hat{y}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_K x_{Ki})$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i + \hat{u}_i) = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i + \underbrace{\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i}_{=0}$$ if there is a constant term among regressors.

- Assumptions for OLS

$y_i = x_i'\beta + u_i$ (Finite sample case)

① $E(u_i | x_i) = 0 \iff E(y_i | x_i) = x_i'\beta$

② $\mathbb{X}$ has full rank (No multicollinearity) & $\mathbb{X}'\mathbb{X}$ is invertible.

③ iid sampling.

④ $E(u_i^2 | x_i = x) = \sigma^2(x) = \sigma^2$ for all $x$ in the support of $X$

WTS : $E(\hat{\beta} | X) = \beta$ Under ①, ②, ③

$$\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'[\mathbb{X}\beta + \mathbb{U}] = \underbrace{(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}}_{I}\beta + (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{U}$$

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

$\boxed{\text{Note}}$ $E(A) = \begin{pmatrix} E(a_{11}) & \cdots & E(a_{1n}) \\ \vdots & & \vdots \\ E(a_{m1}) & \cdots & E(a_{mn}) \end{pmatrix}$

$$E(\hat{\beta} | \mathbb{X}) = E(\beta + (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{U} | \mathbb{X}) = \beta + E((\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{U} | \mathbb{X})$$

$$= \beta + (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' E(\mathbb{U} | \mathbb{X})$$ (By each $x_i$ is non-stochastic)

$$= \beta + (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\begin{pmatrix} E(u_1 | \mathbb{X}) \\ \vdots \\ E(u_N | \mathbb{X}) \end{pmatrix} = \beta + (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\begin{pmatrix} E(u_1 | x_1) \\ \vdots \\ E(u_N | x_N) \end{pmatrix} = \beta.$$

$E(\hat{\beta}) = \beta$ ?

$\rightarrow E(\hat{\beta}) = E(E(\hat{\beta}|X)) = E(\beta) = \beta.$

- Conditional Variance of $\hat{\beta}|X$

$$\mathrm{Var}(\hat{\beta}|X) = E\{[\hat{\beta} - \underbrace{E(\hat{\beta}|X)}_{=\beta}][\hat{\beta} - \underbrace{E(\hat{\beta}|X)}_{=\beta}]' \mid X\}$$

$$= E\{[(X'X)^{-1}X'u][(X'X)^{-1}X'u]' \mid X\}$$

$$= E\{(X'X)^{-1}X'uu'X(X'X)^{-1} \mid X\}$$

$$= (X'X)^{-1}X' \cdot E\{uu' \mid X\} X(X'X)^{-1}$$

$$E(uu'|X) = \begin{pmatrix} E(u_1u_1|X) & E(u_1u_2|X) & \cdots & E(u_1u_N|X) \\ \vdots & & \ddots & \vdots \\ E(u_Nu_1|X) & \cdots & & E(u_Nu_N|X) \end{pmatrix}$$

$u(u_Nu_1|X) = E(u_N|X)E(u_1|X) = 0$ (By assumption ①)

In the same manner, $u(u_su_t|X) = 0 \ \forall s,\ s \neq t.$

$$= \begin{pmatrix} E(u_1^2|X) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & & 0 \\ 0 & 0 & \cdots & E(u_N^2|X) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \Omega$$

(omega matrix)

(By assumption ④ $: E(u_i^2|X_i = x_i) = \sigma^2(x_i) = \sigma^2$)

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

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

Therefore, $\mathrm{Var}(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$

More generally,

$$V(\hat{\beta}|\mathbb{X}) = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' E(uu'|\mathbb{X})\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}$$

$$= (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' \underbrace{\begin{pmatrix} \sigma^2(x_1) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & & \vdots \\ 0 & \cdots & 0 & \sigma^2(x_N) \end{pmatrix}}_{=\Omega} \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}$$

$$= \boxed{(\mathbb{X}'\mathbb{X})^{-1}} \underset{K\times N}{\mathbb{X}'} \underset{N\times N}{\Omega} \underset{N\times K}{\mathbb{X}} \boxed{(\mathbb{X}'\mathbb{X})^{-1}} = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i x_i' \sigma^2(x_i) \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}$$

WTS: $\mathbb{X}'\mathbb{X} = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}$

$$\underset{K\times N\ N\times K}{\mathbb{X}'\mathbb{X}} = \underset{K\times N}{\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{N1} \\ x_{12} & & & \vdots \\ \vdots & & & \vdots \\ x_{1K} & \cdots & & x_{NK} \end{pmatrix}} \underset{N\times K}{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & & & \\ \vdots & & & \\ x_{N1} & \cdots & & x_{NK} \end{pmatrix}}$$

$$= \underset{K\times K}{\begin{pmatrix} x_{11}^2 + x_{21}^2 + \cdots + x_{N1}^2 & x_{11}x_{12} + x_{21}x_{22} + \cdots + x_{N1}x_{N2} & \cdots & x_{11}x_{1K} + \cdots + x_{N1}x_{NK} \\ x_{12}x_{11} + x_{22}x_{21} + \cdots + x_{N2}x_{N1} & & & \\ \vdots & & \ddots & \vdots \\ x_{1K}x_{11} + \cdots + x_{NK}x_{N1} & \cdots & & x_{1K}^2 + \cdots + x_{NK}^2 \end{pmatrix}}$$

Define $\underset{K\times 1}{x_i} := \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{pmatrix}$ Then, $\underset{K\times 1\ 1\times K}{x_1 x_1'} = \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1K} \end{pmatrix} (x_{11}\ x_{12} \cdots x_{1K}) = \begin{pmatrix} x_{11}^2 & x_{11}x_{12} & \cdots & x_{11}x_{1K} \\ \vdots & & \ddots & \vdots \\ x_{1K}x_{11} & \cdots & & x_{1K}^2 \end{pmatrix}$

$x_2 x_2' = \begin{pmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2K} \end{pmatrix} (x_{21}\ x_{22} \cdots x_{2K}) = \begin{pmatrix} x_{21}^2 & x_{21}x_{22} & \cdots & x_{21}x_{2K} \\ \vdots & & \ddots & \vdots \\ x_{2K}x_{21} & \cdots & & x_{2K}^2 \end{pmatrix}$

$\vdots$

Hence, we can know $x_i x_i'$ is the ith term of summation form of each row and column in $\mathbb{X}'\mathbb{X}$.

$$= x_1 x_1' + x_2 x_2' + \cdots + x_N x_N' = \sum_{i=1}^{N} x_i x_i' \quad \text{where } x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{pmatrix}. \quad ▨$$

WTS: $\mathbb{X}'\Omega\mathbb{X} = \sum_{i=1}^{N} x_i x_i' \sigma^2(x_i)$

$$\mathbb{X}'\Omega\mathbb{X} = \underbrace{\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{N1} \\ x_{12} & & & \vdots \\ \vdots & & & \\ x_{1K} & x_{2K} & \cdots & x_{NK} \end{pmatrix}}_{K\times N} \underbrace{\begin{pmatrix} \sigma^2(x_1) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & \sigma^2(x_N) \end{pmatrix}}_{N\times N} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & & & \vdots \\ \vdots & & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2(x_1)x_{11} & \sigma^2(x_2)x_{21} & \cdots & \sigma^2(x_N)x_{N1} \\ \sigma^2(x_1)x_{12} & \cdots & & \sigma^2(x_N)x_{N2} \\ \vdots & & \ddots & \vdots \\ \sigma^2(x_1)x_{1K} & \sigma^2(x_2)x_{2K} & \cdots & \sigma^2(x_N)x_{NK} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & \ddots & & \vdots \\ \vdots & & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2(x_1)x_{11}^2 + \sigma^2(x_2)x_{21}^2 + \cdots + \sigma^2(x_N)x_{N1}^2 & \cdots & \sigma^2(x_1)x_{11}x_{1K} + \sigma^2(x_2)x_{21}x_{2K} + \cdots + \sigma^2(x_N)x_{N1}x_{NK} \\ \vdots & \ddots & \vdots \\ \sigma^2(x_1)x_{1K}x_{11} + \cdots + \sigma^2(x_N)x_{NK}x_{N1} & \cdots & \sigma^2(x_1)x_{1K}^2 + \sigma^2(x_2)x_{2K}^2 + \cdots + \sigma^2(x_N)x_{NK}^2 \end{pmatrix}$$

Define $x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{pmatrix}$ Then $x_1 x_1' \sigma^2(x_1) = \begin{pmatrix} \sigma^2(x_1)x_{11}^2 & \sigma^2(x_1)x_{11}x_{12} & \cdots & \sigma^2(x_1)x_{11}x_{1K} \\ \vdots & & \ddots & \vdots \\ \sigma^2(x_1)x_{1K}x_{11} & & \cdots & \sigma^2(x_1)x_{1K}^2 \end{pmatrix}$

In the same manner, $x_N x_N' \sigma^2(x_N) = \begin{pmatrix} \sigma^2(x_N)x_{N1}^2 & \sigma^2(x_N)x_{N1}x_{N2} & \cdots & \sigma^2(x_N)x_{N1}x_{NK} \\ \vdots & & \ddots & \vdots \\ \sigma^2(x_N)x_{NK}x_{N1} & & \cdots & \sigma^2(x_N)x_{NK}^2 \end{pmatrix}$

Hence, we can know $x_i x_i' \sigma^2(x_i)$ is ith term of summation form of each low and column in $\mathbb{X}'\Omega\mathbb{X}$.

$$= x_1 x_1' \sigma^2(x_1) + x_2 x_2' \sigma^2(x_2) + \cdots + x_N x_N' \sigma^2(x_N) = \sum_{i=1}^{N} x_i x_i' \sigma^2(x_i) \quad \text{where } x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iK} \end{pmatrix}.$$

$\blacksquare$