< 522A Econometrics, 1/14 > class 2

⓪ Review

· Conditional Mean Function : $E(y|x) = \int_{-\infty}^{\infty} y \cdot f(y|x)\, dy$ (y: scalar, x: vector)

$$\begin{aligned}
\cdot\ Var(Y|X=x) &= E\{(Y - E(Y|X=x))^2 \mid X=x\} \\
&= E\{(Y^2 - 2Y \cdot E(Y|X=x) + [E(Y|X=x)]^2) \mid X=x\} \\
&= E\{Y^2|X=x\} - 2E\{Y \cdot E(Y|X=x) \mid X=x\} + E\{[E(Y|X=x)]^2 \mid X=x\} \\
&= E\{Y^2|X=x\} - 2E(Y|X=x) \cdot E(E(Y|X=x)|X=x) + [E(Y|X=x)]^2 \cdot \underbrace{E(1|X=x)}_{=1} \\
&= E\{Y^2|X=x\} - [E(Y|X=x)]^2
\end{aligned}$$

$$\begin{aligned}
\cdot\ Var\,Y &= E(Y - E(Y))^2 = E(Y - E(Y|X) + E(Y|X) - E(Y))^2 \\
&= E\{[Y - E(Y|X)]^2 + [E(Y|X) - E(Y)]^2 + 2[Y - E(Y|X)][E(Y|X) - E(Y)]\} \\
&= \underbrace{E\{[Y - E(Y|X)]^2\}}_{①} + \underbrace{E\{[E(Y|X) - E(Y)]^2\}}_{②} + \underbrace{2E\{[Y - E(Y|X)][E(Y|X) - E(Y)]\}}_{③}
\end{aligned}$$

① $E\{[Y - E(Y|X)]^2\} = E[E\{[Y - E(Y|X)]^2|X\}] = E[Var(Y|X)]$

By $E(Y) = E[E(Y|X)]$

② $E\{[E(Y|X) - E(Y)]^2\} = E\{[E(Y|X) - E(E(Y|X))]^2\} = Var(E(Y|X))$

③ $2E\{[Y - E(Y|X)][E(Y|X) - E(Y)]\} = 2E[E\{[Y - E(Y|X)][E(Y|X) - E(Y)]|X\}]$

전체에 $E(☆) = E(E(☆|X))$

$= 2E\underbrace{[E\{[Y - E(Y|X)]|X\}]}_{=0} \cdot (E(Y|X) - E(Y)) = 0$

$(\because E\{[Y - E(Y|X)]|X\} = E(Y|X) - E(E(Y|X)|X)$
$= E(Y|X) - E(Y|X) \cdot \underbrace{E(1|X)}_{=1} = 0$

$= E[Var(Y|X)] + Var[E(Y|X)]$

· Conditional Average Treatment Effect given $X_2 = x_2$ of changing X_1 from x_1 to x_1'

If w & X_1 are independent given X_2. "assumption" ⊛

$E(Y \mid X_1 = x_1', X_2 = x_2) - E(Y \mid X_1 = x_1, X_2 = x_2)$ ← By ⊛, we can drop ✓ and ✓.

$= E(Y(x_1', x_2, w) \mid X_2 = x_2) - E(Y(x_1, x_2, w) \mid X_2 = x_2)$

$= E[Y(x_1', x_2, w) - Y(x_1, x_2, w) \mid X_2 = x_2]$

⊙ Conditional Expectation (continued)

If $Y(\underbrace{x_1, x_2}_{\text{indexed}}, \underbrace{w}_{\text{random}})$ & X_1 are independent given $X_2 = x_2$ for any x_1, x_2 in the support of (X_1, X_2)

then $E\{Y \mid X_1 = x_1', X_2 = x_2\} - E\{Y \mid X_1 = x_1, X_2 = x_2\}$

$= E\{Y(x_1', x_2, w) - Y(x_1, x_2, w) \mid X_2 = x_2\}$

★ Role of Conditioning

① Make $Y(X_1, X_2, w)$ & X_1 independent given X_2.

② To define parameter of interest

- Example of ① JPE : Rosenzweig & Wolpin

$Y(\text{number of children}, x_2, w)$

⇑

labor force participation decision 0 or 1
Not participation / participation.

ex) $X_2 = \{\text{age, income}, \dots\}$ "parameter of interest group"

$= E\{Y \mid X_1 = 1, X_2 = x_2\} - E\{Y \mid X_1 = 0, X_2 = x_2\}$

ex) • More specific case : ∃ twins or not as the first time child.

$E\{Y \mid X_1 = 1, X_2 = x_2, \text{First time child} = 1\} - E\{Y \mid X_1 = 0, X_2 = x_2, \text{First time child} = 1\}$

. conditioning on age, race is important for reason ①.

If $X_2 = (X_{21}, X_{22})$

$\uparrow$ group of interest — Conditioning variables to secure ①

$$E\{Y(x_1', x_2, w) - Y(x_1, x_2, w) \mid X_{21} = x_{21}\}$$

$$= E\{E(Y(x_1', x_2, w) - Y(x_1, x_2, w) \mid X_2) \mid X_{21} = x_{21}\}$$

: Use distribution of $X_{22} \mid X_{21} = x_{21}$ to integrate out X_{22}

⊙ OLS

- $E(Y|X) := m(x)$, $m(x) = \beta_0 + \beta_1 r_1(x) + \cdots + \beta_K r_K(x)$, $r_j(x), j=1 \ldots K$ are known functions

where $r(x) = \begin{bmatrix} 1 \\ r_1(x) \\ \vdots \\ r_K(x) \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}$ $\beta_0, \beta_1, \ldots, \beta_K$ are known fcns.

ex)

Let $X_1 = x_1$, $X_2 = x_2$.

$$m(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

$r_0(x) = 1$

$r_1(x) = x_1$ $\quad r_3(x) = x_1^2$

$r_2(x) = x_2$ $\quad r_4(x) = x_2^2$ $\quad r_5(x) = x_1 x_2$.

It seems to be six regressors, but two.

$$m(x_1, x_2) = E\{Y \mid \underbrace{X_1 = x_1, X_2 = x_2}_{\#\text{ of conditioning : } 2}\} = [r_0(x), \ldots, r_5(x)] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_5 \end{bmatrix}$$

Not six.

- Often written as $\underline{E\{Y \mid X = x\} = x'\beta}$

$m(x) = x'\beta$, $x = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_K \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}_{(K+1)\times 1}$

"1" is included in this vector but we usually say x is random.

"K+1" is a kind of cumbersome. Therefore, ⇓

- We allow x to include a constant term and keep $\beta \in \mathbb{R}^K$.

- $Y = X'\beta + u$, $\boxed{E(Y|X=x) = x'\beta \iff E\{u|X=x\}=0}$ ⊛

Once $Y = X'\beta + u$ model is maintained

$\Rightarrow$ Assumption for ⊛ ① which X to use

② How to put together $r(x)$

c.f.) In the non-parameter world, we can eliminate these assumptions.

- How to estimate β ?

$$\min_{g(\cdot)} E\{[Y - g(X)]^2\} = E\{[Y - E(Y|X)]^2\} + E\{\underline{[E(Y|X) - g(X)]^2}\}$$

$\Uparrow$ predict Y.

If $E(Y|X) = g(X)$, it is "0"

If $g(x) = x'b$, $E\{[Y - x'b]^2\}$

$$E\{[Y - x'b]^2\} = E\{[Y - \underbrace{E(Y|X)}_{= x'\beta}]^2\} + E\{\underbrace{[E(Y|X) - x'b]^2}_{= [x'(\beta - b)]^2}\}$$

We use sample analog : $\min_{b \in \mathbb{R}^K} \frac{1}{N}\sum_{i=1}^{N} (y_i - x_i'b)^2$

$\Rightarrow$ Sampling of (X, Y) results in (x_i, y_i) for $i = 1, \dots, N$.

The solution to this problem is called the Ordinary Least Squares estimator. (OLS)

- OLS estimator for a matrix form

$$\mathbb{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad \underset{N\times K}{\mathbb{X}} = \begin{pmatrix} x_1' \\ \vdots \\ x_N' \end{pmatrix}$$

$N \times 1$

$\underset{K\times 1}{x_i}$ is a vector

$\Downarrow$

x_i' is $1 \times K$

The objective function is $(\mathbb{Y} - \mathbb{X}b)'(\mathbb{Y} - \mathbb{X}b)$

F.O.C : $-\mathbb{X}'(\mathbb{Y} - \mathbb{X}b) = 0$

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

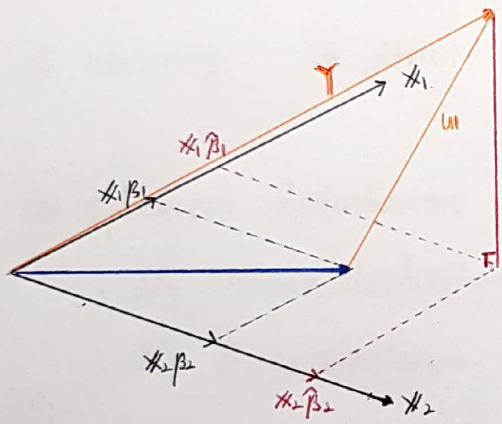
when $\mathbb{X}$ has full rank $\iff \mathbb{X}'\mathbb{X}$ is invertible.

- Geometry of OLS

$\mathbb{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$, $K=2$, $\mathbb{X}_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \end{bmatrix}$, $\mathbb{X}_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ x_{23} \end{bmatrix}$, $\mathbb{U} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$, $\mathbb{X} := [\mathbb{X}_1 \ \mathbb{X}_2]$

$$\mathbb{Y} = \mathbb{X}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathbb{U} = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} x_{11}\beta_1 + x_{21}\beta_2 + u_1 \\ x_{12}\beta_1 + x_{22}\beta_2 + u_2 \\ x_{13}\beta_1 + x_{23}\beta_2 + u_3 \end{bmatrix}$$

$$= \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \end{bmatrix}\beta_1 + \begin{bmatrix} x_{21} \\ x_{22} \\ x_{23} \end{bmatrix}\beta_2 + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \mathbb{X}_1\beta_1 + \mathbb{X}_2\beta_2 + \mathbb{U}$$



$(\mathbb{Y} - \mathbb{X}b)'(\mathbb{Y} - \mathbb{X}b)$: Uclidean distance

$\hat{\mathbb{U}}$ & $\mathbb{X}_1$ are orthogonal
$\hat{\mathbb{U}}$ & $\mathbb{X}_2$ are "
) Thus, $\mathbb{X}_1'\hat{\mathbb{U}} = 0$, $\mathbb{X}_2'\hat{\mathbb{U}} = 0$

F.O.C : $\mathbb{X}'(\mathbb{Y} - \mathbb{X}\hat{\beta}) = 0$

$\mathbb{X} = [\mathbb{X}_1 \vdots \cdots \vdots \mathbb{X}_K]$, $\mathbb{X}' = \begin{bmatrix} \mathbb{X}_1' \\ \vdots \\ \mathbb{X}_K' \end{bmatrix}$ $\qquad \mathbb{X}'\hat{\mathbb{U}} = \begin{bmatrix} \mathbb{X}_1'\hat{\mathbb{U}} \\ \vdots \\ \mathbb{X}_K\hat{\mathbb{U}} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

- Alternative motivation for the OLS

$E\{u|x\} = 0$

$E(xu) = 0 \iff \boxed{\begin{matrix} E(x_1 u) = 0 \\ \vdots \\ E(x_K u) = 0 \end{matrix}} \iff \begin{matrix} \frac{1}{N}\sum_{i=1}^{N} x_{1i}(\underbrace{y_i - x_i'b}_{=u}) = 0 \\ \vdots \\ \frac{1}{N}\sum_{i=1}^{N} x_{Ki}(y_i - x_i'b) = 0 \end{matrix}$

** $\qquad$ (Sample analog of moment conditions)

⇓

Method of Moment motivation for the OLS

c.f) Decomposition of variance

(Note) $\frac{1}{N}\sum_{i=1}^{N} x_i$: sample mean

$\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^2$: sample variance.

SST : Total sum of squares : $\sum_{i=1}^{N}(y_i - \bar{y})^2$

SSR : Regression sum of squares $\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2$

SSE : Error sum of squares $\sum_{i=1}^{N}(y_i - \hat{y}_i)$

Given ① $E(u_i)=0$ and ② $E(x_i u_i)=0$, $SST = SSR + SSE$.

<Basic OLS assumptions.>

$\underset{N\times 1}{\mathbb{Y}} = \underset{N\times K}{\mathbb{X}}\,\underset{K\times 1}{\beta} + \underset{N\times 1}{\mathbb{U}}$ $\quad \left(\underset{1\times 1}{y_i} = \underset{1\times K}{x_i'}\,\underset{K\times 1}{\beta} + \underset{1\times 1}{u_i},\ i=1,\dots,N \right)$

① linearity ② strict exogeneity : $E(u_i|X)=0 \Leftrightarrow E(X'u)=0$

$\parallel$ orthogonality

$E(x_i u_i)=0$ & $E(u_i)=0$

③ spherical disturbances

- No Heteroskedasticity : $E(u_i^2|X)=\sigma^2,\ \forall i$
- No Autocorrelation : $E(u_t u_s|X)=0,\ \forall t\neq s$.

④ $\mathbb{X}$ has full column rank → $rank(\mathbb{X})=K$

↓

No multicollinearity ($\Rightarrow$ guarantees that $\mathbb{X}'\mathbb{X}$ is invertible)

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

proof) $\min SSE = \min (\mathbb{Y}-\mathbb{X}\beta)'(\mathbb{Y}-\mathbb{X}\beta) = \mathbb{Y}'\mathbb{Y} - \mathbb{Y}'\mathbb{X}\beta - (\mathbb{X}\beta)'\mathbb{Y} + (\mathbb{X}\beta)'\mathbb{X}\beta$

$= \mathbb{Y}'\mathbb{Y} - \mathbb{Y}'\mathbb{X}\beta - \beta'\mathbb{X}'\mathbb{Y} + \beta'\mathbb{X}'\mathbb{X}\beta$

↓ Take derivative w.r.t β

$$\frac{\partial SSE}{\partial \beta} = 0 - (\mathbb{Y}'\mathbb{X})' - \mathbb{X}'\mathbb{Y} + (\beta'\mathbb{X}'\mathbb{X})' + \mathbb{X}'\mathbb{X}\beta = -2\mathbb{X}'\mathbb{Y} + 2\mathbb{X}'\mathbb{X}\beta$$

Thus, $-2\mathbb{X}'\mathbb{Y} + 2\mathbb{X}'\mathbb{X}\hat{\beta} = 0$ i.e., $\mathbb{X}'\mathbb{X}\hat{\beta} = \mathbb{X}'\mathbb{Y}$

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

(Also, $\mathbb{X}'\mathbb{Y} - \mathbb{X}'\mathbb{X}\hat{\beta} = 0 \Leftrightarrow \mathbb{X}'(\mathbb{Y} - \mathbb{X}\hat{\beta}) = 0$) ∎