Lecture 3

Regressor sub-vector.

$X = (X_1, X_2)$.

$y = x'\beta + e = x_1'\beta + x_2'\beta + e$. $E(xe) = 0$.

$$Q_{xx} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = \begin{pmatrix} E x_1' x_1 & E x_1' x_2 \\ E x_2' x_1 & E x_2' x_2 \end{pmatrix}$$

$$Q_{xy} = \begin{pmatrix} Q_{1y} \\ Q_{2y} \end{pmatrix} = \begin{pmatrix} E x_1 y \\ E x_2 y \end{pmatrix}.$$

The formula of partitioned matrix inverse.

$$Q_{xx}^{-1} = \begin{pmatrix} Q_{11} & Q_{21} \\ Q_{21} & Q_{22} \end{pmatrix}^{-1}$$

$$= \begin{bmatrix} \underline{(Q_{11} - Q_{12} Q_{22}^{-1} Q_{21})^{-1}} & -\underline{(Q_{11} - Q_{12} Q_{22}^{-1} Q_{21})^{-1}} Q_{12} Q_{22}^{-1} \\ -\underline{(Q_{22} - Q_{21} Q_{11}^{-1} Q_{12})^{-1}} Q_{21} Q_{11}^{-1} & \underline{(Q_{22} - Q_{21} Q_{11}^{-1} Q_{12})^{-1}} \end{bmatrix}$$

$$= \begin{bmatrix} Q_{11\cdot 2}^{-1} & -Q_{11\cdot 2}^{-1} Q_{12} Q_{22}^{-1} \\ Q_{22\cdot 1}^{-1} Q_{21} Q_{11}^{-1} & -Q_{22\cdot 1}^{-1} \end{bmatrix}$$

therefore $\beta = \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} = \begin{pmatrix} \quad \end{pmatrix}\begin{pmatrix}Q_{1y}\\ Q_{2y}\end{pmatrix}$

$$= \begin{pmatrix} Q_{11\cdot2}^{-1}(Q_{1y} - Q_{12}Q_{22}^{-1}Q_{2y}) \\ Q_{22\cdot1}^{-1}(Q_{2y}-Q_{21}Q_{11}^{-1}Q_{1y})\end{pmatrix}$$

$$= \begin{pmatrix}Q_{11\cdot2}^{-1}Q_{1y\cdot2}\\ Q_{22\cdot1}^{-1}Q_{2y\cdot1}\end{pmatrix}$$

~~why is this so~~

$$Q_{1y\cdot2} = E(x_1 y) - E(X_1X_2)E(X_2X_2)^{-1}E(X_2 y).$$

Let $\dim(X_1) = 1$. why is this useful?

consider $X_1 = X_2'\gamma + u_1$, $E(X_2u_1) = 0$.

$\gamma_2 = Q_{22}^{-1}Q_{21}$.

$E(u_1 y) = \cancel{E y(X_1 - X_2' E(X_2X_2)^{-1}E(X_2 X_1))}$ $= E(X_1 - X_2'E(X_2X_2)^{-1}E(X_2X_1)y)$

y is scalar, can be moved.

$= E(X_1y) - E(yX_2)E(X_2X_2)^{-1}E(X_2X_1)$

$= Q_{1y\cdot2}$.

$E(u_1^2) = \cancel{E((X_1 - X_2'E(X_2X_2)^{-1}E(X_2X_1))^2)}$

$\cancel{= EX_1^2}$ $= E(X_1 - X_2'\gamma)^2 = EX_1^2 - 2E(X_1X_2)\gamma + \gamma'E(X_2X_2)\gamma$

$= EX_1^2 - 2E(X_1X_2)E(X_2X_2)E(X_2X_1) + E(X_1X_2)(E(X_2X_2))^{-1}\cancel{E(X_2X_2)}\cancel{E(X_2X_2)^{-1}}E(X_2X_1)$

$= EX_1^2 - E(X_1X_2)EX_2X_2\,EX_2X_1 = Q_{11}$

So $\beta_1 = (E u_1^2)^{-1} E(u_1 y)$

the coefficient β equals the projection coefficient of y on u_1. u_1 is the error from X_1 on other variables X_2, u_1 is the component of X_1 ~~which is~~ not linearly explained by X_2. "Pure effect" of X_1 net of the linear effect of X_2.

Omitted ~~Bia~~ Variable Bias (OVB)

Think about the case with no constant → Standard the scale

if we omitted X_2. (e.g. X_2 is unobservable).

$$y = X_1'\gamma + u, \quad E(X_1 u) = 0.$$

The short regression

(cf. $y = X_1'\beta_1 + X_2\beta_2 + e$) the long regression

In general $\gamma_1 \neq \beta_1$.

$$\gamma_1 = E(X_1 X_1')^{-1} E(X_1 y)$$

$$= E(X_1 X_1') E X_1 (X_1'\beta_1 + X_2\beta_2 + e)$$

$$= \cancel{E(X_1 X_1')^{-1} E X_2}\ \beta_1 + \boxed{E(X_1 X_2)^{-1} E(X_2 X_2)\beta_2}$$

OVB

Unless $E(X_1 X_2) = 0$ (uncorrelated) or $\beta_2 = 0$. should not be added

X_1 has a constant, X_2 has no constant, $E(\text{constant}, X_2) = 0$ means X_2 has zero mean.

6

~~The linear predictor is also the line~~

Sometimes we can sign the OVB. This is an important inference in practice.

$y \leftarrow$ wage. $X_1 \leftarrow$ education $X_2 \leftarrow$ IQ.

$\beta_2 > 0$, $E(X_1, X_2) > 0$, so that.

OVB > 0. ~~we~~ $\gamma_1 > \beta_1$

We have an upward bias in using the short regression.

The linear ~~regres~~ predictor is the best linear approximation to the CEF.

$$\min_{\beta} E\left(m(x) - x\beta\right)^2.$$

irrelevant to β.

best predictor. $\beta = E(XX')^{-1}E(X, y)$.

$$\cancel{E(m(x) - x\beta) = E((m(x) - y)^2 + E(y - x\beta)^2}$$

$$\cancel{+2E\left[(m(x) - y)(y - x\beta)\right]}$$

~~by LIE, E[E(m(x) - y~~

~~$\beta = E(\ldots$~~ by formula, $\beta^* = \left(E(XX')\right)^{-1} E(X\, m(x))$

$$= E(XX')^{-1} E(Xy).$$

Because $E(Xy) = E(X(m(x) + e)) = E(m(x)\, X)$.

Causal effect

$y = h(x_1, x_2, u)$

scalar, parameter of interest.

control variable

all other variables.

"Causal effect of x_1" is the chg in y due to the change in x_1 holding ~~all the other~~ other factors x_2 and u constant.

$$C(x_1, x_2, u) = \frac{\partial}{\partial x_1} h(x_1, x_2, u)$$

if x_1 is differentiable.

$$C(x_1, x_2, u) = h(x_1 = 1, x_2, u) - h(x_1 = 0, x_2, u)$$

if x_1 is binary.

~~Either can be written~~ as $\nabla_1 h(x_1, x_2, u)$

In the potential outcome framework, discuss x_1 binary.

$y(1)$, $y(0)$. $C(x_2, u) = y(1) - y(0)$.

unobservable.

$E(y(1) - y(0))$ ATE.

$E((y(1) - y(0)) \mid x_1 = 1)$. ATET

more generally ACE $= \int \nabla_1 h(x_1, x_2, u) f(u \mid x_1, x_2)\, du$. derivative average across the marginal.

Does the $m(x)$ ~~~~ implies the ATE / ACE) average causal effect?

or is $\nabla_1 m(x_1, x_2) = ACE$?

Because

$$m(x_1, x_2) = \cancel{E(h(X_1} \; E(y \mid X_1, X_2)$$

$$= E(\cancel{y}\, e + h(x_1, x_2, u) \mid X_1, X_2)$$

$$= E(h(x_1, x_2, u) \mid X_1, X_2)$$

$$= \int h(x_1, x_2, u) f(u \mid x_1, x_2)\, du$$

aver $h(x_1, x_2, u)$ over the conditional dist of u.

two terms in has u. assume integrable of intg and diff.

However, $\nabla_1 m(x_1, x_2) = \int \nabla_1 h(x_1, x_2, u) f(u \mid x_1, x_2)\, du$

$$+ \int h(x_1, x_2, u) \nabla_1 f(u \mid x_1, x_2)\, du$$

$\frac{\partial}{\partial x_1} f(u_1 \mid x_1, x_2)$

$$= ACE + \int h(x_1, x_2, u) \nabla_1 f(u \mid x_2, x_2)\, du$$

CLA: Conditional on X_2, ~~~~, the r.v.s X_1 and u are stat. indep.

So $\nabla_1 f(u \mid x_1, x_2) = \nabla_1 f(u \mid x_2) = 0$

Because x_1 doesn't enter, must be 0. ~~indp~~ irrelevant to x_1

9