Regression with one r.h.s. variable

1. (a) Show that solutions to the ordinary least squares (OLS) problem

$$\underset{b_0,b_1}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

are given by

$$\hat{\beta}_1 := \frac{\widehat{\operatorname{cov}}(X,Y)}{\widehat{\operatorname{var}}(X)} \qquad \hat{\beta}_0 := \overline{Y} - \hat{\beta}_1 \overline{X}.$$

[Hint: show that the estimators solve appropriate analogues of the normal equations.]

(b) Explain why the residuals $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ satisfy $\sum_{i=1}^{n} \hat{u}_i = 0$ and $\sum_{i=1}^{n} X_i \hat{u}_i = 0$.

(c) What is the relationship between $\hat{\beta}_1$ and the sample correlation between Y_i and X_i?

$$\hat{\beta}_1 = \frac{\widehat{Cov}(X,Y)}{\widehat{Var}(X)} \qquad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

We are minimising the sum of squares:

$$C = \min \sum_i (Y_i - b_0 - b_1 X_i)^2$$

By the FOCs,

$$\frac{\partial C}{\partial b_0} = \sum (Y_i - b_0 - b_1 X_i) = 0 \qquad (1)$$

$$\frac{\partial C}{\partial b_1} = \sum X_i (Y_i - b_0 - b_1 X_i) = 0$$

From (1),

$$\sum (Y_i - b_0 - b_1 X_i) = 0$$

$$\Rightarrow \sum Y_i - b_1 \sum X_i = n b_0$$

$$\Rightarrow \tfrac{1}{n} \sum Y_i - b_1 \tfrac{1}{n} \sum X_i = b_0$$

$$\Rightarrow b_0 = \overline{Y} - b_1 \overline{X} \qquad (3)$$

From (2):

$$\sum (X_i)(Y_i - b_0 - b_1 X_i) = 0$$

$$\Rightarrow \sum X_i Y_i - b_0 \sum X_i - b_1 \sum X_i^2 = 0$$

$$\Rightarrow \sum X_i Y_i - (\overline{Y} - b_1 \overline{X}) \sum X_i = b_1 \sum X_i^2$$

☆

$$\hat{Var}(X) = \Sigma(X_i - \bar{X})^2$$
$$= \Sigma(X_i^2 - 2X_i\bar{X} + \bar{X}^2)$$
$$= \Sigma X_i^2 - 2\Sigma X_i \bar{X} + \Sigma \bar{X}^2$$
$$= \Sigma X_i^2 - 2n\bar{X}^2 + \Sigma \bar{X}^2$$
$$= \Sigma X_i^2 - 2n\bar{X}^2 + n\bar{X}^2$$
$$= \Sigma X_i^2 - n\bar{X}^2$$

$$\hat{Cov}(A,B) = \Sigma(X_i - \bar{X})(Y_i - \bar{Y})$$
$$= \Sigma(X_iY_i - X_i\bar{Y} - \bar{X}Y_i + \bar{X}\bar{Y})$$
$$= \Sigma(X_iY_i) - \bar{Y}\Sigma X_i - \bar{X}\Sigma Y_i + \Sigma\bar{X}\bar{Y}$$
$$= \Sigma X_iY_i - n\bar{Y}\bar{X} - n\bar{X}\bar{Y} + \Sigma\bar{X}\bar{Y}.$$
$$= \Sigma X_iY_i - n\bar{X}\bar{Y}$$

$\Rightarrow$ The FOCs give the optimal levels of b_0^* and b_1^* that solve the least mean squared problem, which can also be obtained by some function of the sample variance / covariance and sample mean.

b) Explain why the residuals $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ satisfy $\sum^n \hat{u}_i = 0$ and $\hat{\sum} \hat{u}_i X_i = 0$.

$\rightarrow$ We've shown that $\hat{\beta}_0$ and $\hat{\beta}_1$ are the solutions to the least-squared problem. That is to say,

$$\Sigma Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i = 0 \quad \text{from FOC \#1}$$

$\Rightarrow \Sigma \hat{u}_i = 0.$ Similarly,

$$\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)X_i = 0 \quad \text{from FOC \#2}.$$

$\Rightarrow \Sigma \hat{u}_i X_i = 0.$ $\square$

Given that they are the solutions to the FOCs, they satisfy by construction these two equalities.

(c) What is the relationship between β̂1 and the sample correlation between Yi and

$$\hat{\beta}_1 = \frac{Cov(\bar{Y}, \bar{X})}{Var(\bar{X})}$$

$$\text{Sample correlation} = \frac{S_{XY}}{S_X S_Y} \qquad \frac{\frac{1}{n-1} \sum (X_i - \bar{X})(Y_i - \bar{Y})}{\cancel{\frac{1}{n-1}} \sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}}$$

$$= \frac{Cov(\bar{X}, \bar{Y})}{\sqrt{Var(\bar{X}) Var(\bar{Y})}}$$

$$\Rightarrow \quad \hat{\beta}_1 = \sqrt{\frac{Var(\bar{Y})}{Var(\bar{X})}} \cdot \frac{S_{XY}}{S_X \cdot S_Y}$$

I thought it would but can't get it to work when solving the FOC

2. Consider the (population) regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

where $\mathbb{E}u_i = 0$ and $\mathbb{E}X_i u_i = 0$. Assuming $\beta_1 \neq 0$, we can rewrite this as

$$X_i = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} Y_i - \frac{1}{\beta_1} u_i.$$

Does it follow that $1/\beta_1$ gives the coefficient on Y_i in a population linear regression of X_i on Y_i? Explain.

→ I think it must do...

We can similarly do the minimisation of least squares?

Let $X_i = \gamma_0 + \gamma_1 Y_i - \gamma_1 u_i$

$C = \min \sum (X_i - \gamma_0 - \gamma_1 Y_i)^2$

FOCs are still

$$\frac{\partial C}{\partial \gamma_0} = 2\sum(X_i - \gamma_0 - \gamma_1 Y_i) = 0$$

$$\frac{\partial C}{\partial \gamma_1} = 2\sum(X_i - \gamma_0 - \gamma_1 Y_i)\gamma_1 = 0$$

$\Rightarrow \sum X_i - n\gamma_0 - \gamma_1 \sum Y_i = 0$

$\Rightarrow \sum X_i - \gamma_1 \sum Y_i = n\gamma_0$

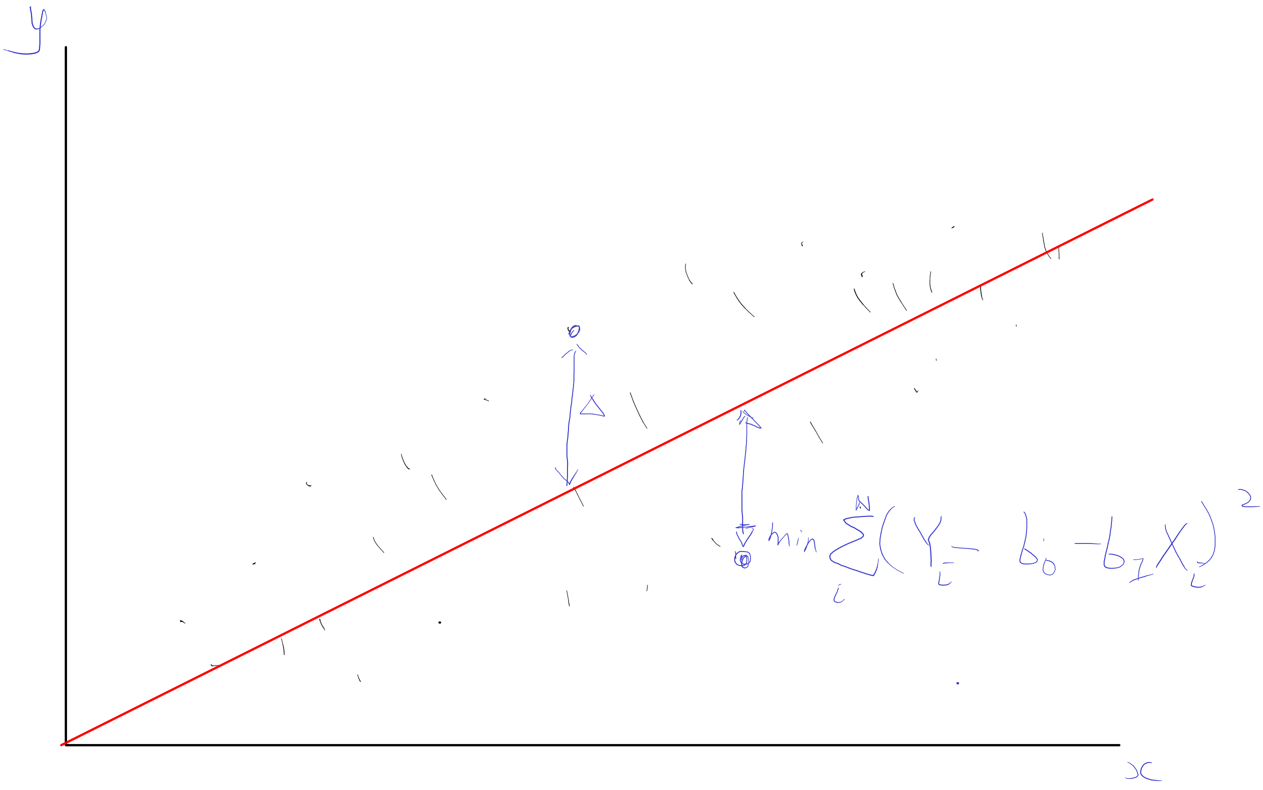
$\Rightarrow \gamma_0 = \bar{X} - \gamma_1 \bar{Y}$ $\qquad \beta_0 = \bar{Y} - \beta_1 \bar{X}$

Similar derivation gives

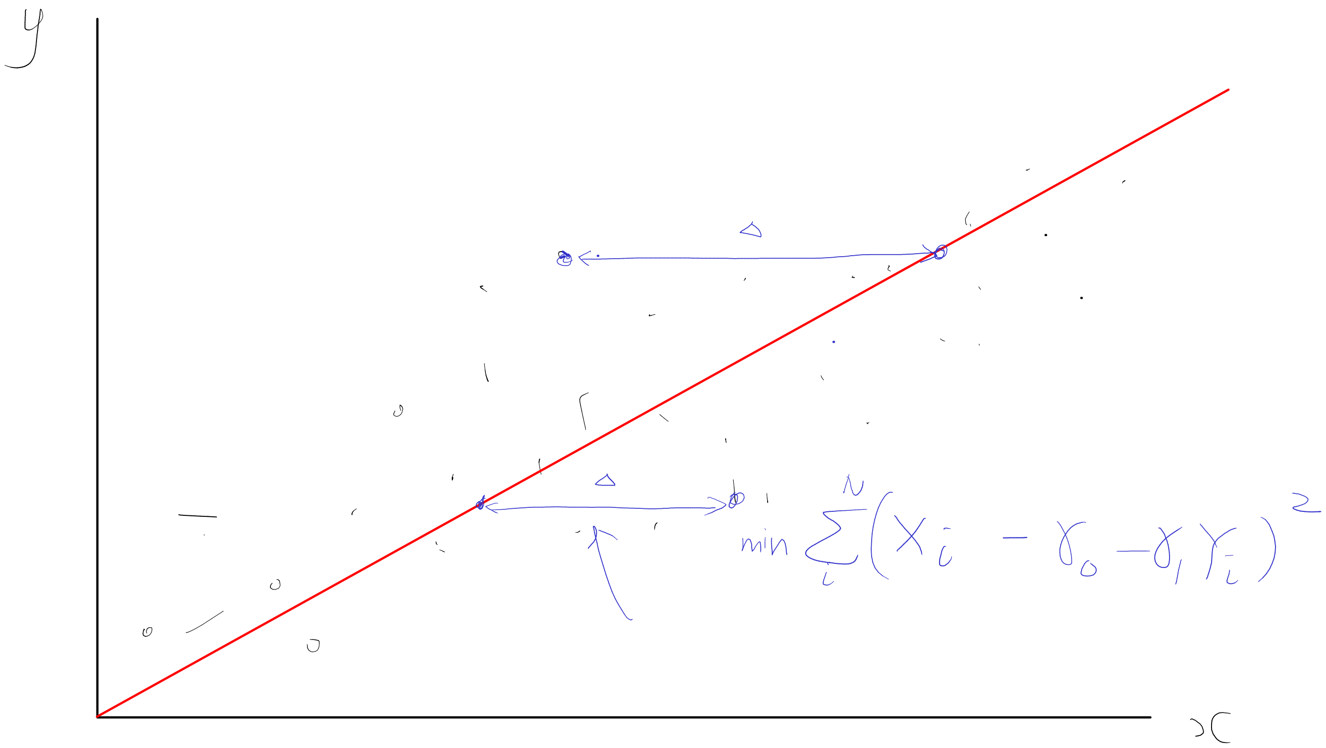
$$\hat{\gamma}_1 = \frac{\hat{Cov}(X,Y)}{\hat{Var}(Y)} \qquad \hat{\beta}_1 = \frac{\hat{Cov}(X,Y)}{\hat{Var}(X)}$$

Unless $\hat{Var}(Y) = \hat{Var}(X)$, they won't be equal.

Intuition as follows:



Minimising the sum of mean squares distance over the y-axis.



Minimising the sum of mean squares distance over the x-axis.

Regression with multiple r.h.s. variables

3. Suppose that

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

where $\mathbb{E}u_i = 0$, $\mathbb{E}X_{1i}u_i = 0$ and $\mathbb{E}X_{2i}u_i = 0$ (assumption OR). Show that $(\beta_0, \beta_1, \beta_2)$ solve the population linear regression problem

$$\min_{b_0,b_1,b_2} \mathbb{E}(Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2.$$

[Hint: there is no need to derive explicit expressions for the solution to this problem.]

What's the idea here?

→ The causal model parameters solve the pop LR problem

→ we know from the FOCs of the pop LR problem that

$$\mathbb{E}(Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}) X_\ell = 0$$

where $X_0 = 1$.

→ From the causal model and taking expectations on both sides we have

$$\mathbb{E}(Y_i) = \beta_0 + \beta_1 \mathbb{E}X_{1i} + \beta_2 \mathbb{E}X_{2i} + \mathbb{E}u_i$$

$$\Rightarrow \mathbb{E}[Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i}] = \mathbb{E}u_i$$

Similarly,

$$\mathbb{E}(Y_i X_\ell) = \beta_0 X_\ell + \beta_1 \mathbb{E}X_{1i} X_\ell + \beta_2 \mathbb{E}X_{2i} X_\ell + \mathbb{E}u_i X_\ell \quad \text{for } \ell = \{1, 2\}$$

$$\Rightarrow \mathbb{E}[Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i}] X_\ell = \mathbb{E}u_i X_\ell$$

We are given that $\mathbb{E} u_i = 0$, $\mathbb{E} u_i X_\ell = 0$

for $\ell = \{1, 2\}$ and so

$$\mathbb{E}\left[Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i}\right] X_\ell = \mathbb{E} u_i X_\ell$$

for $\ell = 0, 1, 2$

4. Consider the model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i.$$

Show that the OLS estimator $\hat{\beta}_1$ satisfies

$$\hat{\beta}_1 = \frac{\hat{\text{cov}}(Y, \tilde{X}_1^{\wedge})}{\hat{\text{var}}(\tilde{X}_1^{\wedge})}$$

where $\tilde{X}_{1i}^{\wedge}$ denotes the residual from an OLS regression of X_{1i} on X_{2i}.

$$X_{1i} = \gamma_0 + \gamma_1 X_{2i} + \tilde{X}_{1i}$$

Substitute this into the causal model:

$$Y_i = \beta_0 + \beta_1(\gamma_0 + \gamma_1 X_{2i} + \tilde{X}_{1i}) + \beta_2 X_{2i} + u_i$$

$$= \underbrace{\beta_0 + \beta_1\gamma_0}_{\pi_0} + \underbrace{(\beta_1\gamma_1 + \beta_2)}_{\pi_2} X_{2i} + \beta_1 \tilde{X}_{1i} + u_i$$

$$= \pi_0 + \pi_2 X_{2i} + \beta_1 \tilde{X}_{1i} + u_i$$

$$Cov(Y, \tilde{X}_1) = cov(\overset{=0}{\cancel{\pi_0}} + \beta_1 \tilde{X}_1 + \overset{=0 \text{ (by construction)}}{\cancel{\pi_2 X_{2i}}} + u_i, \tilde{X}_1)$$

$$= \beta_1 Var(\hat{X}_1) + Cov(u_i, \tilde{X}_1)$$

$$= \beta_1 Var(\hat{X}_1) + Cov(u_i, X_1 - \cancel{\delta_0} - \gamma_1 X_2)$$

$$+ \underbrace{Cov(u_1, X_2)}_{0R} + \underbrace{Cov(X_1, u_1)}_{0R}$$

$$\Rightarrow \beta_1 = \frac{Cov(Y, \tilde{X}_1)}{Var(\tilde{X}_1)}$$

The sample analogue is the same, so

$$\hat{\beta}_1 = \frac{\hat{Cov}(Y, \hat{\tilde{X}}_1)}{\hat{Var}(\tilde{X}_1)}$$

Conditional expectations

5. By using the decomposition

$$Y = \mathbb{E}[Y \mid X] + \varepsilon$$

where $\varepsilon := Y - \mathbb{E}[Y \mid X]$, prove the *law of total variance*:

$$\begin{aligned} \operatorname{var} Y &= \operatorname{var}(\mathbb{E}[Y \mid X]) + \mathbb{E}\varepsilon^2 \\ &= \operatorname{var}(\mathbb{E}[Y \mid X]) + \mathbb{E}[\operatorname{var}(Y \mid X)] \end{aligned}$$

where $\operatorname{var}(Y \mid X) := \mathbb{E}[(Y - \mathbb{E}[Y \mid X])^2 \mid X]$.

$$\mathbb{E}(\varepsilon) = \mathbb{E}(\mathbb{E}(\varepsilon \mid X)) = \mathbb{E}(0) = 0$$

$$Y = \mathbb{E}[Y \mid X] + \varepsilon \quad \text{where } \mathbb{E}(\varepsilon \mid X) = 0$$
$$\mathbb{E}(\varepsilon) = 0$$

$$\operatorname{var}(Y) = \operatorname{var}(\mathbb{E}[Y \mid X] + \varepsilon)$$
$$= \operatorname{Var}(\mathbb{E}[Y \mid X]) + 2\operatorname{Cov}(\mathbb{E}[Y \mid X], \varepsilon) + \operatorname{Var}(\varepsilon)$$

As $\operatorname{Cov}(\mathbb{E}[Y \mid X]) = \mathbb{E}(\mathbb{E}[Y \mid X] \cdot \varepsilon) - \cancel{\mathbb{E}(\mathbb{E}[Y \mid X]) \cdot \underbrace{\mathbb{E}(\varepsilon)}_{=0}}$

$$\overset{LIE}{=} \mathbb{E}[\mathbb{E}(\mathbb{E}[Y \mid X] \cdot \varepsilon \mid X)]$$
$$\overset{\text{conditioning}}{=} \mathbb{E}[\mathbb{E}[Y \mid X] \cdot \underbrace{\mathbb{E}(\varepsilon \mid X)}_{=0}]$$
$$= 0.$$

$$\Rightarrow \operatorname{var}(Y) = \operatorname{Var}(\mathbb{E}[Y \mid X]) + \operatorname{Var}(\varepsilon)$$
$$= \operatorname{Var}(\mathbb{E}[Y \mid X]) + \mathbb{E}(\varepsilon^2) - \cancel{\underbrace{\mathbb{E}(\varepsilon)^2}_{=0}}$$

Shown.

6. Show that the parameters of a population linear regression of Y on X also minimise

$$\mathbb{E}\{\mathbb{E}[Y \mid X] - (b_0 + b_1 X)\}^2$$

with respect to (b_0, b_1). [Hint: first read the proof given in Appendix C.2 of the notes.]

① We know that a pop LR of Y on X:

gives us b_0^*, b_1^* such that

$$b_0^* = \mathbb{E}[Y] - b_1^* \mathbb{E}(X) \quad (1a)$$

$$b_1^* = \frac{Cov(X, Y)}{Var(X)} \quad (2a)$$

② We also know that minimising the least squared problem,

$\underset{b_0, b_1}{\operatorname{argmin}} \mathbb{E}\{(\mathbb{E}[Y|X] - b_0 + b_1 X)^2\}$ gives us

b_0^*, b_1^* such that

$$b_0^* = \mathbb{E}(\mathbb{E}[Y|X]) - b_1^* \mathbb{E}(X) \quad (1b)$$

$$b_1^* = \frac{Cov(X, \mathbb{E}[Y|X])}{Var(X)} \quad (2b)$$

③ If we can prove (1a) = (1b) AND (2a) = (2b),

we are done.

⇒ The CEF decomposition allows us to write Y as the following:

$$Y = \mathbb{E}[Y|X] + \varepsilon \qquad \text{where } \mathbb{E}(\varepsilon|X) = 0.$$

and by LIE $\mathbb{E}(\varepsilon) = 0$.

$\Rightarrow$ By ~~substituting~~ $Y = \mathbb{E}[Y|X] + \varepsilon$ into (1a) and (2a) we obtain the following:

$$b_0^* = \mathbb{E}\big[\mathbb{E}[Y|X] + \varepsilon\big] - b_1^* \mathbb{E}(X) \qquad (1c)$$

$$b_1^* = \frac{Cov(X, \mathbb{E}[Y|X] + \varepsilon)}{Var(X)} \qquad (2c)$$

First working with 1(c) we have

$$b_0^* \overset{\text{linearity}}{=} \mathbb{E}\big[\mathbb{E}[Y|X]\big] + \underbrace{\cancel{\mathbb{E}(\varepsilon)}}_{\mathbb{E}(\varepsilon)=0} - b_1^* \mathbb{E}(X)$$

$$\overset{\mathbb{E}(\varepsilon)=0}{=} \mathbb{E}\big[\mathbb{E}[Y|X]\big] - b_1^* \mathbb{E}(X)$$

which equals (1b), and

$$b_1^* = \frac{Cov(X, \mathbb{E}[Y|X] + \varepsilon)}{Var(X)}$$

$$= \frac{\underbrace{\cancel{Cov(X, \varepsilon)}}_{=0^*} + Cov(\mathbb{E}[Y|X], X)}{Var(X)}$$

$$= \frac{Cov(\mathbb{E}[Y|X], X)}{Var(X)}$$

* As $Cov(X, \varepsilon) = \mathbb{E}(\varepsilon X) - \cancel{\mathbb{E}(\varepsilon)}\mathbb{E}(X)$
$= \mathbb{E}(\varepsilon X)$
$= \mathbb{E}(\mathbb{E}(\varepsilon X | X))$
$= \mathbb{E}(X \underbrace{\cancel{\mathbb{E}(\varepsilon|X)}}_{=0}) = 0$

which is equivalent to (2b), so we are done.

2019 Q5

5. Consider the model

$$Y_i = \beta_0 + \beta_{1i} X_i + u_i,$$

where β_{1i}, the causal effect of X_i on Y_i, is itself a random variable (it varies across individuals).

(a) [20%] Suppose that u_i and β_{1i} are both mean independent of X_i. Show that a population linear regression of Y_i on X_i (and a constant) would recover $\mathbb{E}\beta_{1i}$. [Hint: what is $\mathbb{E}[Y_i \mid X_i]$?]

(b) [10%] Give a brief interpretation of $\mathbb{E}\beta_{1i}$.

STRATEGY: show the CEF coincides with the pop LR.

① take the conditional expectation on both sides.

$$\mathbb{E}[Y_i \mid X_i] = \beta_0 + \mathbb{E}[\beta_{1i} X_i \mid X_i] + \mathbb{E}(u_i \mid X_i)$$

$$\overset{MI}{=} \beta_0 + \mathbb{E}\beta_{1i} \cdot X_i + u_i \qquad = 0$$

$\Rightarrow$ The best predictor of Y_i with X_i alone is

$$\beta_0 + \mathbb{E}\beta_{1i} \cdot X_i + u_i$$

② The pop LR solves the least squared problem

$$\underset{b_0, b_1}{\operatorname{argmin}} \ \mathbb{E}(Y - b_0 - b_1 X)^2$$

and is the best linear predictor of Y_i with X_i alone.

$$b_1^* = \frac{Cov(X_i, Y_i)}{Var(X_i)}$$

③ Because the CEF is linear, the best predictor is the best linear predictor and thus the pop LR's solutions coincide with the CEF. Thus

$$b_1^* = \mathbb{E}\beta_{1i} = \frac{Cov(X_i, Y_i)}{Var(X_i)}$$

QED □

Suppose now that u_i and β_{1i} are not necessarily mean independent of X_i, but there is an 'instrument' Z_i which is related to X_i by the equation

$$X_i = \pi_0 + \pi_{1i} Z_i + v_i,$$

and is such that u_i, v_i, β_{1i} and π_{1i} are independent (not merely mean independent) of Z_i. Let β_{IV} denote the coefficient in a population two-stage least squares regression of Y_i on X_i, using Z_i as an instrument. That is, β_{IV} is obtained by the following procedure:

i. X_i is regressed on Z_i and a constant (in the population), to obtain fitted values $X_i^* := \delta_0 + \delta_1 Z_i$.

ii. Y_i is regressed on X_i^* and a constant (in the population); β_{IV} is the coefficient on X_i^* in this regression.

Now answer the following questions.

(c) [20%] Show that

$$\beta_{\text{IV}} = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(X_i, Z_i)}. \tag{1}$$

(d) [20%] Using (1), show that

$$\beta_{\text{IV}} = \mathbb{E}\left\{\beta_{1i} \frac{\pi_{1i}}{\mathbb{E}\pi_{1i}}\right\}. \tag{2}$$

$$X_i^* := \delta_0 + \delta_1 Z_i$$

$$Y_i = \beta_0 + \beta_{1i} X_i + u_i$$

$$Y_i = \beta_0 + \beta_{IV} X_i^* + u_i$$

$$\beta_{IV} = \frac{\text{Cov}(Y_i, X_i^*)}{\text{var}(X_i^*)}$$

$$= \frac{\text{Cov}(Y_i, \delta_0 + \delta_1 Z_i)}{\text{var}(\delta_0 + \delta_1 Z_i)}$$

$$= \frac{\overbrace{\cancel{Cov(Y_i, \delta_0)}}^{=0} + \delta_1 Cov(Y_i, Z_i)}{\cancel{Var(\delta_0)} + \delta_1^2 Var(Z_i) + \underbrace{\cancel{2cov(\delta_0, \delta_1 Z_i)}}_{=0}}$$

$$= \frac{\delta_1 Cov(Y_i, Z_i)}{\delta_1^2 Var(Z_i)}$$

$$= \frac{Cov(Y_i, Z_i)}{Var(Z_i)} \cdot \frac{1}{\delta_1}$$

From the OLS regression of Z_i on X_i

$$X_i = \pi_0 + \pi_1 Z_i + v_i$$

$$\delta_1 = \frac{Cov(X_i, Z_i)}{Var(Z_i)}$$

$$\Rightarrow \frac{Cov(Y_i, Z_i)}{\cancel{Var(Z_i)}} \cdot \frac{\cancel{Var}(Z_i)}{Cov(X_i, Z_i)}$$

$$= \frac{Cov(Y_i, Z_i)}{Cov(X_i, Z_i)}$$

$\square$

(d) [20%] Using (1), show that

$$\beta_{\text{IV}} = \mathbb{E}\left\{\beta_{1i}\frac{\pi_{1i}}{\mathbb{E}\pi_{1i}}\right\}. \tag{2}$$

$$X_i = \pi_0 + \pi_{1i} Z_i + v_i .$$

$$\beta_{IV} = \frac{Cov(Y_i, Z_i)}{Cov(X_i, Z_i)}$$

$$= \frac{Cov(Y, Z_i)}{Var(Z_i)} \cdot \frac{Var(Z_i)}{Cov(X_i, Z_i)}$$

We have the causal model $Y_i = \beta_0 + \beta_{1i} X_i + u_i$

Substituting to get reduced form:

$$Y_i = \beta_0 + \beta_{1i}\pi_0 + \underbrace{\beta_{1i}\pi_{1i}}\, Z_i + \beta_{1i} v_i + u_i$$

Taking conditional expectations on both sides,

$$\mathbb{E}[Y_i | Z_i] = \mathbb{E}[\beta_0 + \beta_{1i}\pi_0 + \beta_{1i}\pi_i Z_i | Z_i] + \mathbb{E}[\overbrace{\beta_{1i} v_i + u_i}^{e} | Z_i]$$

By independence ($\beta_0 \perp\!\!\!\perp Z_i$, β_{1i}, v_j $u_i \perp\!\!\!\perp Z_i$),

$$= \mathbb{E}[\beta_0 + \beta_{1i}\pi_0 + e] + \mathbb{E}\,\beta_{1i}\pi_i \cdot Z_i$$

which is linear in Y_i.

$\Rightarrow$ A pop LR of Y_i on Z_i recovers $b_1^* = \dfrac{Cov(Z_i, Y_i)}{Var(Z_i)}$

From part I we know that if the CEF is linear, then the pop LR and the CEF coincide:

Given that it is, $\mathbb{E}\beta_{1i}\pi_1 = b_1^*$, thus

$$\mathbb{E}\beta_{1i}\pi_1 = \frac{Cov(Y_i, Z_i)}{Var(Z_i)}$$

A similar argument holds for the regression of X_i on Z_i:

The pop LR minimises

$$\underset{\pi_0, \pi_1}{\operatorname{argmin}} (X_i - \pi_0 - \pi_1 Z_i)^2$$

and it can be shown

$$\pi_1 = \frac{Cov(X_i, Z_i)}{Var(Z_i)}$$

Taking conditional expectations on both sides,

$$\mathbb{E}(X_i | Z_i) = \mathbb{E}(\pi_0 + \pi_{1i} Z_i + v_i | Z_i)$$

As $\pi_{1i} \perp\!\!\!\perp Z_i$, $v_i \perp\!\!\!\perp Z_i$

$$= \pi_0 + \mathbb{E}\pi_{1i} \cdot Z_i + v_i$$

which is linear in Z_i

$\Rightarrow$ best predictor of X_i is linear

$\Rightarrow$ CEF coincides with LR

$$\Rightarrow \mathbb{E}\pi_{1i} = \frac{Cov(X_i, Z_i)}{Var(Z_i)}$$