Condition for an unbiased estimate

- simple regression:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

We can estimate in $\underline{\text{any}}$ case!

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

But will estimates be good?

- On lecture: $E(\hat{\beta}_1 | X) = \beta_1$

$$E(E(\hat{\beta}_1 | X)) = \underbrace{E\hat{\beta}_1 = \beta_1}_{\text{unbiased estimate}}$$

if $\boxed{E(u_i | X) = 0}$

our best guess about u_i (build on information about X) is zero.

① weaker ② weaker

u and x are independent $\xleftarrow{\times}$ $E(u_i | X_i) = 0$ $\rightarrow$ $E(u) = 0$

$E(u_i | X_i) = 0 \rightarrow \text{corr}(u, X) = 0$

③

weaker

we are here

(1)

	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
weather u	-1	1	0
mood X	0	0	1

$$E(u|X) = \begin{cases} 0 & X=0 \\ 0 & X=1 \end{cases}$$

but

$$E(X|u) = \begin{cases} 0 & u=\{-1,1\} \\ 1 & u=0 \end{cases}$$

$\rightarrow$ not independent

(2)

	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
weather u	-1	0	1
mood X	0	1	1

$$E(u|X) = \begin{cases} -1, & X=0 \\ \frac{1}{2}, & X=1 \end{cases}$$

but

$E(u)=0$

(3) $X \sim U(-1,1)$

$u = X^2$

$\Rightarrow$ • $\operatorname{cov}(X,u) = E(Xu) - EXEu$

• $E(Xu) = E(X^3) \stackrel{=}{}$

$$= \int_{-1}^{1} x^3 \cdot \frac{1}{2} dx = \frac{1}{2} \cdot \frac{1}{4} x^4 \Big|_{-1}^{1} = 0$$

• $EX = 0 \Rightarrow \operatorname{cov}(X,u) = 0$

But $E(u|x) = E(x^2|x) = x^2 \neq 0$ (usually)

so
$$\boxed{\begin{aligned} &corr(x,u)=0 \not\Rightarrow E(u|x)=0 \\ &E(u)=0 \not\Rightarrow E(u|x)=0 \\ &E(u|x)=0 \not\Rightarrow \text{independ.} \end{aligned}}$$

$$\begin{cases} cov(u,x) = E(ux) - EuEx = E(ux) = 0 \\ E(u|x)=0 \Rightarrow E(E(u|x)) = Eu = 0 \end{cases}$$

$$E(ux) = E(E(ux|x)) = E(x\underbrace{E(u|x)}_{=0}) = 0$$

$$\Rightarrow \boxed{E(u|x)=0 \Rightarrow cov(u,x)=0}$$

usefull $\boxed{corr(x,u) \neq 0 \Rightarrow E(u|x) \neq 0}$

OK, $\forall i = 1, \dots n \quad E(u_i|x_i) = 0$

Why it is so important?

Sometimes we can build a model where
this assumption goes wrong :(

This assumption goes wrong :(

Task 1: Missing var.

salary, educ, profession

real life: $\underset{\text{salary}}{y_i} = \beta_0 + \beta_1 \underset{\text{educ}}{X_{i1}} + \beta_2 \underset{\text{profession}}{X_{i2}} + u_i$

our model $y_i = \beta_0 + \beta_1 X_{i1} + V_i$

$E(u_i | X_i) = 0$ Is $E(\hat{\beta}_1) = \beta_1$?

$cov(X_{i1}, X_{i2}) \neq 0$

our estimation of the slope $\hat{\beta}_1 = \dfrac{\sum (X_{i1} - \bar{X}_1)(y_i - \bar{y})}{\sum (X_{i1} - \bar{X}_1)^2} =$

$$= \frac{\sum (X_{1i} - \bar{X}_1) y_i}{\sum X_{i1}(X_{i1} - \bar{X}_1)} =$$

$$= \frac{\sum (X_{1i} - \bar{X}_1)(\beta_0 + \beta_1 X_{i1} + V_i)}{\sum X_{i1}(X_{i1} - \bar{X}_1)} =$$

$$= \frac{\beta_0 \cdot 0}{\sum X_{i1}(X_{i1} - \bar{X}_1)} + \beta_1 \frac{\sum X_{i1}(X_{i1} - \bar{X}_1)}{\sum X_{i1}(X_{i1} - \bar{X}_1)} + \frac{\sum V_i (X_{i1} - \bar{X}_1)}{\sum (X_{i1} - \bar{X}_1)^2} =$$

$$= \beta_1 + \frac{\sum V_i (X_{i1} - \bar{X}_1)}{\sum (X_{i1} - \bar{X}_1)^2}$$

$$E(\hat{\beta}_1 | X_1) = \beta_1 + \frac{\sum (X_{i1} - \bar{X}_1) E(V_i | X_{i1})}{\sum (X_{i1} - \bar{X}_1)^2} \neq \beta_1 !$$

in fact $V_i = \beta_2 X_{i2} + u_i$

$$E(V_i | X_{i1}) = \beta_2 E(X_{i2} | X_{i1}) + E(u_i | X_{i1}) =$$

$$= \beta_2 E(X_{i2} | X_{i1}) \neq 0$$

because $cov(X_{i1}, X_{i2}) \neq 0 \Rightarrow$

$\Rightarrow$ $\hat{\beta}_1$ is biased in this case

you can only miss regressors which are uncorrelated with yours.

Task2: hands shake

real model: $y_i = \beta_0 + \beta_1 X_i + u_i$

$E(u_i | X_i) = 0$

But observed $\tilde{X}_i = X_i + V_i$

V_i is independent with X_i, u_i ...

so available model $y_i = \beta_0 + \beta_1 \tilde{X}_i + e_i \Rightarrow$

$$\Rightarrow E(\hat{\beta}_1) = \beta_1 + \frac{\sum(\tilde{X}_i - \bar{\tilde{X}}) E(e_i | \tilde{X}_i)}{\sum(\tilde{X}_i - \bar{\tilde{X}})^2}$$

in our task $e_i = \beta_1 X_i - \beta_1 \tilde{X} + u_i =$

$= -\beta_1 \cdot V_i + u_i$

$E(e_i | \tilde{X}_i) = E(-\beta_1 V_i + u_i | \tilde{X}_i) \neq 0$

Because $cov(-\beta_1 V_i + u_i , \tilde{X}_i) =$

$= cov(-\beta_1 V_i + u_i , X_i + V_i) =$

$= -\beta_1 \underbrace{cov(V_i, X_i)}_{=0} - \beta_1 \underbrace{cov(V_i, V_i)}_{\neq 0} +$

$+ \underbrace{cov(u_i, X_i)}_{=0} + \underbrace{cov(u_i, V_i)}_{=0} = -\beta_1 Var V_i$

so shaking hand lead to biased estimate

We discussed $E(\hat{\beta})$ but what about $Var(\hat{\beta})$ - ?

It is also important because we need to calculate the error of our estimation (and to test hypotheses)

(and to test hypotheses)

Task 3 $\quad y_i = \beta_0 + \beta_1 x_i + u_i$

a) $\hat{\beta}_1 = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \dfrac{\sum(x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum(x_i - \bar{x})^2} =$

$$= 0 + \beta_1 \frac{\sum x_i(x_i - \bar{x})}{\sum x_i(x_i - \bar{x})} + \frac{\sum u_i(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} =$$

$$= \beta_1 + \frac{\sum u_i(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}.$$

$$Var(\hat{\beta}_1|X) = 0 + Var\left(\frac{\sum u_i(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} \Big| X\right) =$$

$$= \frac{Var(\sum u_i(x_i - \bar{x})|X)}{[\sum(x_i - \bar{x})^2]^2} \; \textcircled{=}$$

[we need some assumpt. about u_i

$\boxed{Var(u_i|X) = \sigma^2}$]

$$\textcircled{=} \frac{\sum Var(u_i(x_i - \bar{x})|X)}{[\sum(x_i - \bar{x})^2]^2} = \frac{\sum(x_i - \bar{x})^2 Var(u_i|X)}{[\sum(x_i - \bar{x})^2]^2} =$$

$$= \frac{\sigma^2 \sum(x_i - \bar{x})^2}{[\sum(x_i - \bar{x})^2]^2} = \boxed{\frac{\sigma^2}{\sum(x_i - \bar{x})^2} = Var(\hat{\beta}_1|X)}$$

Attention! σ^2 $(u_i \sim F(0, \sigma^2))$

should be known for this formula

8) If we don't know true variance of u_i then we don't know

$Var(\hat{\beta}_1 | x)$:(

what can we do? Estimate it!

$\sigma^2 \longrightarrow \hat{\sigma}^2$

unknown estimation

$$\Rightarrow \quad var(\hat{\beta}_1 | X_1) \longrightarrow \widehat{var}(\hat{\beta}_1 | X_1) = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2}$$

$$\boxed{\hat{\sigma}^2 = \frac{1}{n-2} \sum \hat{u}_i^2 = \frac{RSS}{n-2}}$$

unbiased,

$E(\hat{\sigma}^2) = \sigma^2$

$$\Rightarrow \quad \widehat{var}(\hat{\beta}_1 | x) = \frac{\frac{RSS}{n-2}}{\sum (X_i - \bar{X})^2}$$

(n–k for k regressors)

Task 4 $y_i = \beta_0 + \beta_1 x_i + u_i$

x_i	y_i
1	1
2	1
3	4

$E(u_i|x_i)=0$

$var(u_i|x_i) = 0{,}5$

$cov(u_i, u_j) = 0$

a) $Var(\beta_1)$, $Var(\hat{\beta}_1|X)$, $\widehat{Var}(\hat{\beta}_1|x)$

- $Var(\beta_1) = 0$ ← fixed param.

- $Var(\hat{\beta}_1|x) = \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2} = \frac{0{,}5}{(1-2)^2+0+(3-2)^2} = \frac{0{,}5}{2} = 0{,}25$

- $\widehat{Var}(\hat{\beta}_1|x) = \frac{\frac{RSS}{n-2}}{\Sigma(x_i-\bar{x})^2} = \frac{\Sigma\hat{u}_i^2}{(3-2)\cdot 2}$

$\hat{\beta}_1 = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} = \frac{3}{2}$ (from sem1)

$\hat{\beta}_0 = -1$ (from sem 1)

$\hat{y} = \begin{pmatrix} -1+\frac{3}{2}\cdot 1 \\ -1+3/2\cdot 2 \\ -1+3/2\cdot 3 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 2 \\ 7/2 \end{pmatrix} \Rightarrow \hat{u} = \begin{pmatrix} 1/2 \\ -1 \\ 1/2 \end{pmatrix}$

$\Rightarrow \Sigma\hat{u}_i^2 = \frac{1}{4}+1+\frac{1}{4} = \frac{3}{2} \Rightarrow$

$$\Rightarrow Var(\hat{\beta}_1 | x) = \frac{3/2}{2} = 0{,}75$$