The Simple Regression Model: Functional Form, Expected Values, and Variances

Functional Form

$\hat{\beta}_1 = 0.6$

$y = \beta_0 + \beta_1 x + u$ For $\Delta x = 1$, $\Delta y = 0.6$

y: hrly. wage (\$)

x: educ (yrs.)

$\log(y) = \beta_0 + \beta_1 \log(x) + u$ i.e. $\hat{\beta}_1$

$\hat{\beta}_1 = 0.7$

- Logarithm For $\Delta x = 1\%$

y and x	
$\log(y)$ and x	
$\log(y)$ and $\log(x)$	

$\log(y) = \beta_0 + \beta_1 x + u$

$\downarrow$ natural log

$\hat{\beta}_1 = 0.08$

Approximate effect of $\Delta x = 1$

$\rightarrow \% \Delta y = 100 \hat{\beta}_1$

$= 100 \times 0.08$ (for small $\hat{\beta}_1$)

Exact effect $\rightarrow$

$\% \Delta y = 100[\exp(\hat{\beta}_1) - 1]$

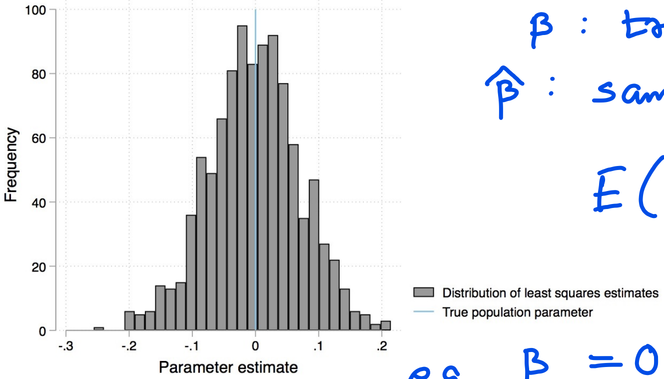
$= 100 [0.083]$

$\Delta y = 0.7\%$

i.e. $\hat{\beta}_1 = \% \Delta y$ [elasticity of y w.r.t. x]

Expected Values of the OLS Estimators

- Under certain assumptions, the OLS estimators are unbiased



Causal Inference: The Mixtape

$\hat{\beta}$

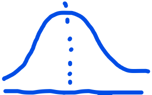
β : true value

$\hat{\beta}$: sample estimates

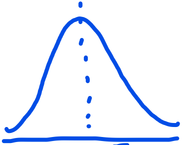
$$E(\hat{\beta}) = \beta$$

e.g. $\beta = 0.5$

$\tilde{\beta}$



0.5



2.5

Expected Values of the OLS Estimators (cont.)

$y = \beta_0 + \beta_1 x + u$

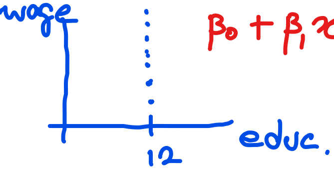
Violation of SLR.2

$Trade = \beta_0 + \beta_1 \, trade\ agreement + u$

only have trade > $10,000 in sample

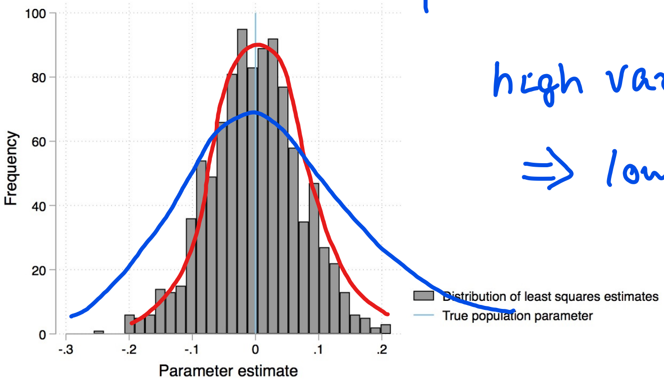
- Assumptions (for unbiasedness)
  - (SLR.1) Linear in Parameters
  - (SLR.2) Random Sampling
  - (SLR.3) Sample Variation in x
    - ★ The value of x - not constant in sample
  - (SLR.4) Zero Conditional Mean
    - ★ $E(u|x) = 0$ → violated e.g. if x and u correlated → x is endogenous; if not violated → x is exogenous
    - ★ $E(y|x) = \beta_0 + \beta_1 x$

violation



Variance of the OLS Estimators

- Variance of an estimator : precision



high variance

$\Rightarrow$ low precision

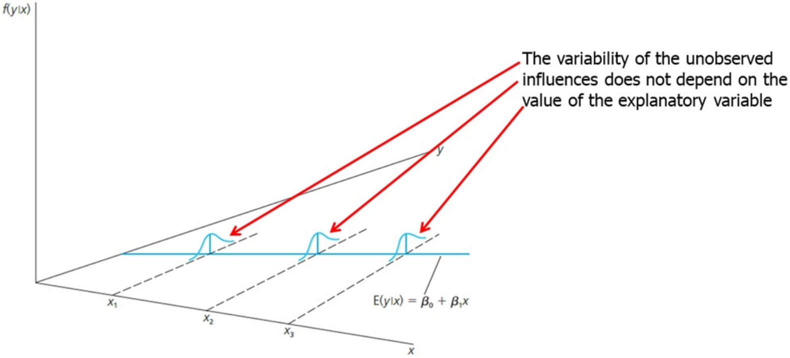
Causal Inference: The Mixtape

Variance of the OLS Estimators (cont.)

- (SLR.5) Homoskedasticity

$$Var(u|x) = Var(u) = \sigma^2$$

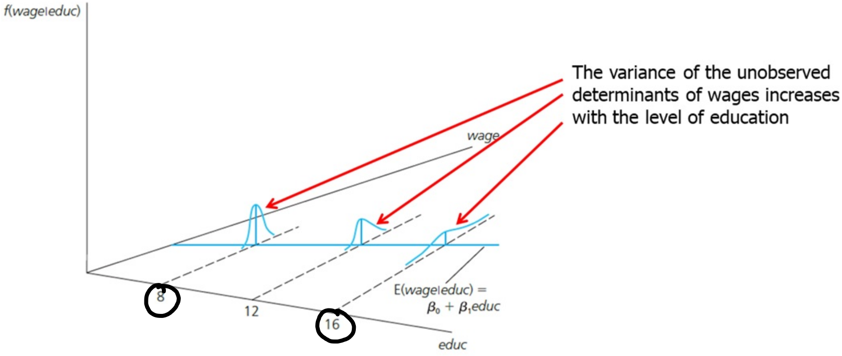
$$Var(y|x) = \sigma^2$$



Variance of the OLS Estimators (cont.)

heteroskedasticity

- Violation of homoskedasticity → heteroskedasticity



Variance of the OLS Estimators (cont.)

- Under SLR.1 to SLR.5

$$Var(\hat{\beta}_1) = \sigma^2 / SST_x$$

$$SST_x = \sum_i (x_i - \bar{x})^2$$

- Standard deviation

$$sd(\hat{\beta}_1) = \frac{\sigma}{\sqrt{SST_x}}$$

$\sigma \rightarrow \sigma$: unknown

Variance of the OLS Estimators (cont.)

$$\hat{u}_i = y_i - \hat{y}_i$$

- Under SLR.1 to SLR.5, $\hat{\sigma}^2$ is an unbiased estimator of σ^2

$$\hat{\sigma}^2 = \frac{SSR}{(n-2)} = \frac{\sum_i \hat{u}_i^2}{(n-2)}$$

- Estimator of σ

$$\hat{\sigma} = \sqrt{\frac{SSR}{(n-2)}}$$

with $\hat{\sigma}$: std. error of regression

- Standard error of $\hat{\beta}_1$

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{SST_x}}$$

Variance of the OLS Estimators (cont.)

y (wage)	x (educ)	$\hat{y}$	$\hat{u}$	$\hat{u}^2$	$(x-\overline{x})^2$
3.1	11	4.498	−1.398	1.955	3.61
3.2	12	5.679	−2.439	5.950	0.81
3	11	4.498	−1.498	2.245	3.61
6	8	0.955	5.045	25.447	24.01
5.3	12	5.679	−0.379	0.144	0.81
8.8	16	10.403	−1.653	2.732	9.61
11	18	12.765	−1.515	2.294	26.01
5	12	5.679	−0.679	0.461	0.81
3.6	12	5.679	−2.079	4.323	0.81
18	17	11.584	6.596	43.510	16.81

$SST_x = 86.9$ $SSR = 89.061$

$\boxed{86.9}$

$\hat{\sigma} = 3.34$ $se(\hat{\beta}_1) = 0.358$

$$\hat{\sigma} = \sqrt{\frac{89.061}{(10-2)}} = 3.34$$

$$se(\hat{\beta}_1) = \frac{3.34}{\sqrt{86.9}}$$

Final Thoughts

- Unlikely that $\hat{\beta}_1$ from a simple regression captures the true effect of x on y
- Need u and x uncorrelated
- Randomized trials help in some cases
  - https://www.youtube.com/watch?v=eGRd8jBdNYg&list=PL-uRhZ_p-BM5ovNRg-G6hDib27OCvcyW8&index=4