Lecture 21 13.04.2021

Econometrics-II

HETEROSKEDASTICITY : $\boxed{\text{SECOND MOMENT PROPERTY}}$

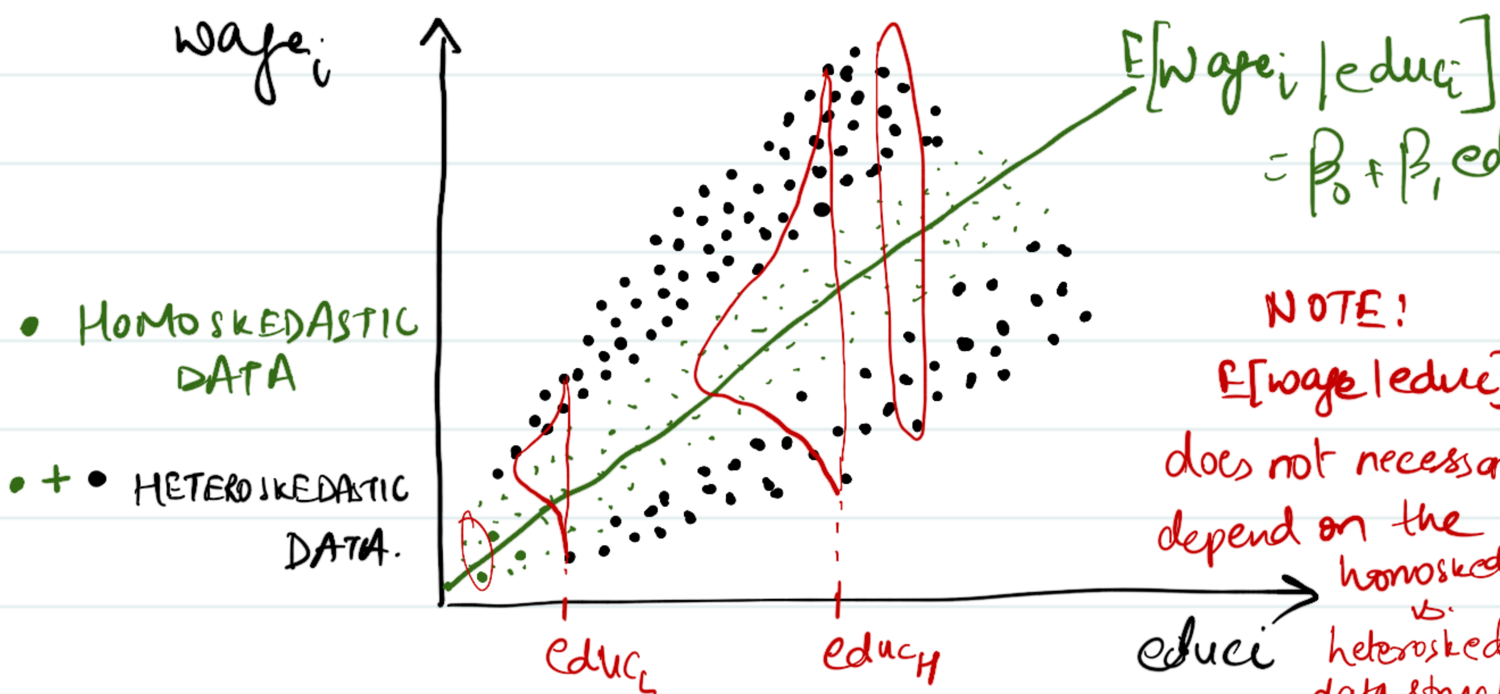
vs. HOMOSKEDASTICITY

→ the regression error has a constant variance across the sample (w.r.t. regressors, x_i's)

$$V(u_i|x_i) = \sigma^2 \quad \forall\, i = 1, 2, \ldots N$$

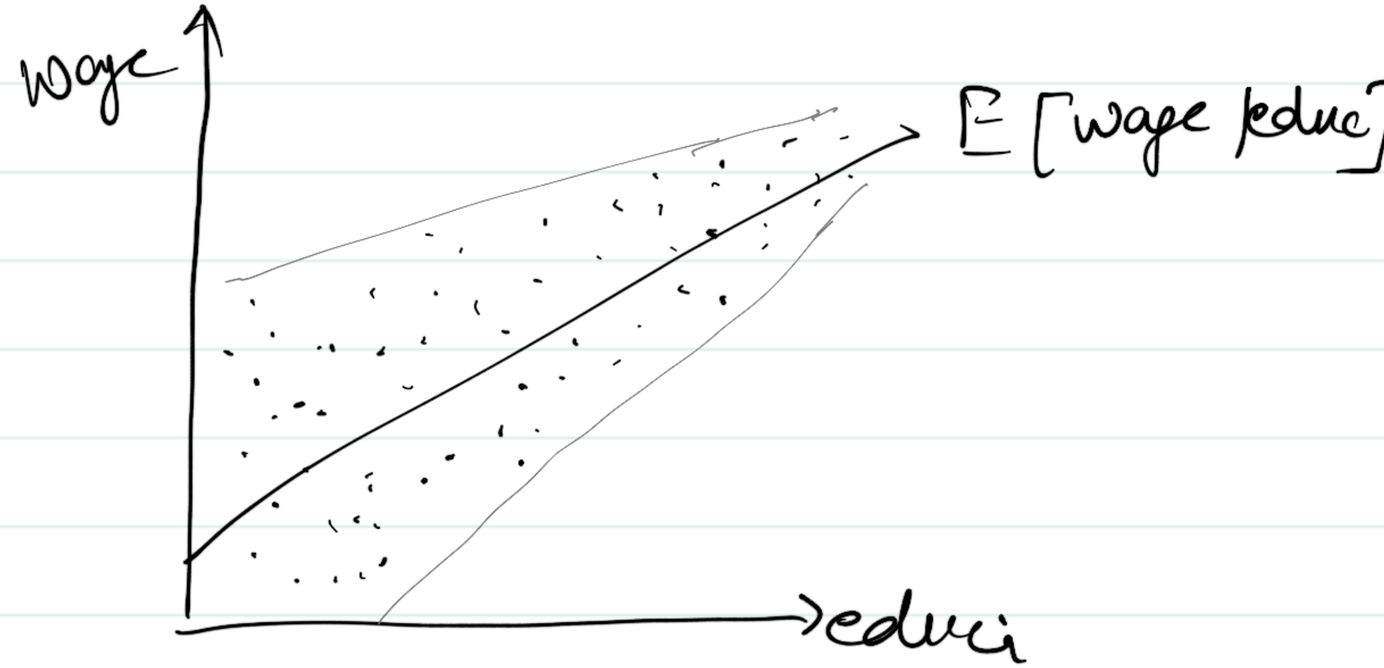
→ $V(u|x_i) = \sigma_i^2$

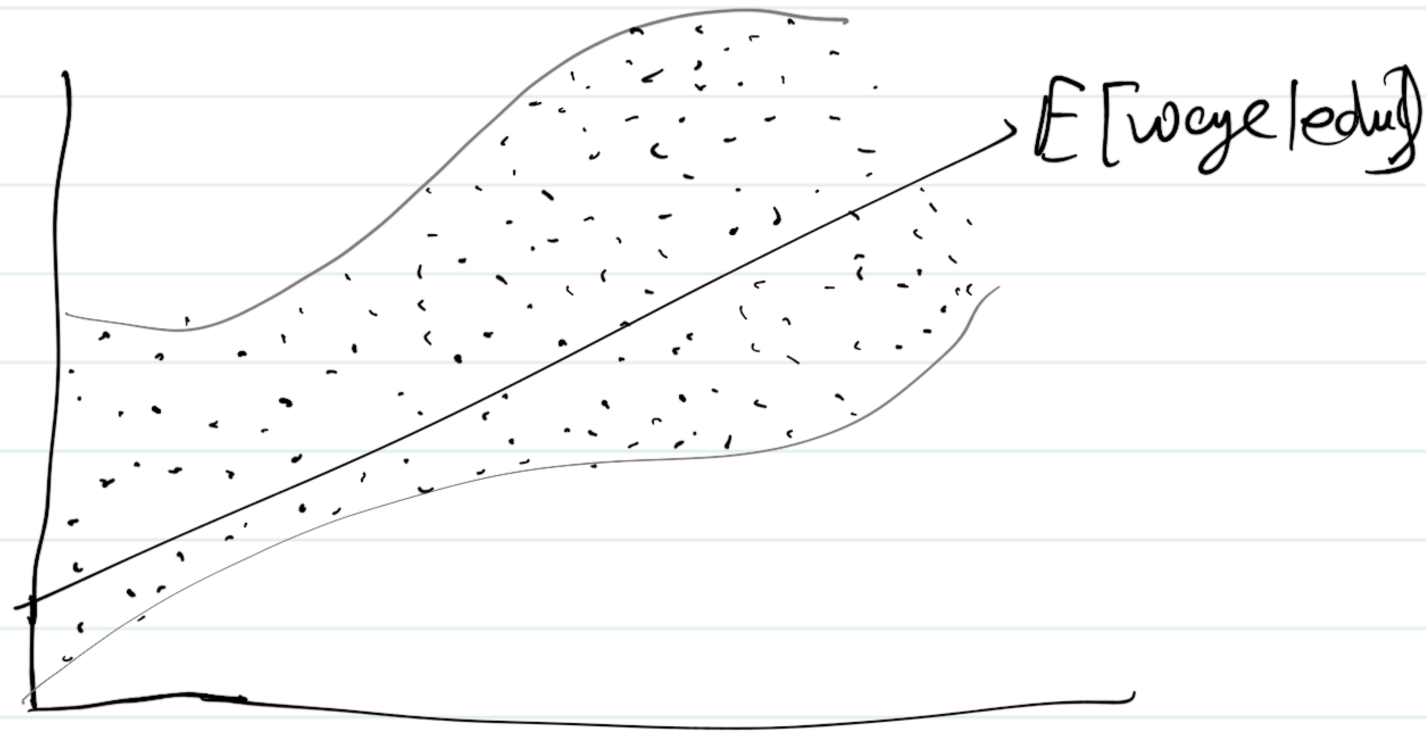
$\boxed{\text{wage vs. education example}}$



NOTE!
$E[wage|educ]$ does not necessarily depend on the homoskedastic vs. heteroskedastic data structure.

Alternative Heteroskedastic structures





Consequence of heteroskedasticity?

The case of simple regression model

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

Heteroskedastic errors: $V(u_i \mid x_i) = \sigma_i^2$

$i = 1, 2, \ldots, N,$

OLS:

$$\hat{\beta}_{1,OLS} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})\, y_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$= \beta_0 \frac{\cancelto{0}{\sum_{i=1}^{N}(x_i - \bar{x})}}{\sum_{i=1}^{N}(x_i - \bar{x})^2} + \beta_1 \cancelto{1}{\frac{\sum_{i=1}^{N}(x_i - \bar{x})\, x_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} + \frac{\sum_{i=1}^{N}(x_i - \bar{x})\, u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$\hat{\beta}_{1,OLS} = \beta_1 + \frac{\sum_{i=1}^{N} (x_i - \bar{x})\, u_i}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Take expectation on both sides w.r.t. x_i's

$$E\left(\hat{\beta}_{1,OLS}\right) = \beta_1 + \frac{\sum_{i=1}^{N} E(u_i x_i)}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

$\therefore$ For $\hat{\beta}_{1OLS}$ to be unbiased (i.e, $E(\hat{\beta}_1) = \beta$)

we require $E(u_i x_i) = 0$ $\rightarrow$ to do with whether we can simulate a ceteris paribus experiment.

But, unbiasedness of $\hat{\beta}_{1OLS}$ has nothing to do with heteroskedasticity.

$$\hat{\beta}_{1,OLS} = \beta_1 + \frac{\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

random variable

$$V(\hat{\beta}_{1,OLS}|x_i) = V\left(\beta_1 + \frac{\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right)$$

const.

interchangeably used the notation $V(\hat{\beta}_{1,OLS})$

$$= \frac{1}{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^2} \sum_{i=1}^{N}(x_i-\bar{x})^2 V(u_i|x)$$

Under heteroskedasticity: $V(u_i|x) = \sigma_i^2$

Hence, $V(\hat{\beta}_{1,OLS}) = \frac{1}{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^2} \sum_{i=1}^{N}(x_i-\bar{x})^2\sigma_i^2$

Under homoskedastic errors:

$$V(\hat{\beta}_{1,OLS}) = \frac{1}{D^{r}} \sum_{i=1}^{N}(x_i-\bar{x})^2 \sigma^2$$

Under homoskedastic errors :-

$$V(\hat{\beta}_{1,OLS}) = \frac{\sigma^2 \cancel{\sum_{i=1}^{N} (x_i - \bar{x})^2}}{\left[\sum_{i=1}^{N} (x_i - \bar{x})^2\right]^{\cancel{2}}}$$

$$= \frac{\sigma^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Under heteroskedastic errors:

$$V(\hat{\beta}_{1,OLS}) = \frac{\sum_{i=1}^{N}\left[(x_i - \bar{x})^2 \sigma_i^2\right]}{\left[\sum_{i=1}^{N} (x_i - \bar{x})^2\right]^2}$$

So, $$V(\hat{\beta}_{1,OLS})\Big|_{HETEROSKEDASTIC} \neq V(\hat{\beta}_{1,OLS})\Big|_{HOMOSKEDASTIC}$$

even when $$E(\hat{\beta}_{1,OLS})\Big|_{HETEROSKEDASTIC} = E(\hat{\beta}_{1,OLS})_{HOMOSKEDASTIC} = \beta_1$$

The impact of heteroskedastic errors shows up while conducting statistical inference.

How?

Say, $H_0 : \beta_1 = 0$

$H_a : \beta_1 \neq 0$

$$t = \frac{\hat{\beta}_{1,ols} - 0}{s.e.(\hat{\beta}_{1,ols})} \sim t_{N-K}$$

different under Hetoskedastic vs homoskedastic errors.

$\Rightarrow t_{\text{Homoskedastic errors}} \neq t_{\text{HETEROSKEDASTIC ERRORS}}$

So Inference, $t \gtrless t_{N-K,\alpha}$ $\Rightarrow$ depend on the error structure.

[depends on error structure]

[comes from t-table so constant]

Aside:

How will we write s.e. under heteroscedastic errors?

$$V(\hat{\beta}_{1,OLS})\Big|_{Homoskedastic} = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

$$s.e.(\hat{\beta}_{1,OLS})\Big|_{\text{—''—}} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}}$$

$$V(\hat{\beta}_{1,OLS})\Big|_{Heteroskedastic} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2\sigma_i^2}{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^2}$$

$$s.e.(\hat{\beta}_{1,OLS})\Big|_{Heteroskedastic} = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2\hat{\sigma}_i^2}{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^2}}$$

Under Heteroskedasticity

OLS is no longer efficient.

$\Rightarrow$ OLS is NOT BLUE

$\Rightarrow$ we can find an estimator with lower variance than OLS.

remedy: estimate β

GENERALIZED LEAST SQUARES ESTIMATORS

$\boxed{\text{GLS estimators}}$

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad i = 1, 2, \dots, N$$

$$V(u_i | x_i) = \sigma_i^2$$

The variance-covariance matrix of regression errors can be written as:

$$\sigma_i = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots\ 0 & \sigma_N^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} \sigma_1^2/\sigma^2 & & & & \\ & \sigma_2^2/\sigma^2 & & 0 & \\ & & \sigma_3^2/\sigma^2 & & \\ & 0 & & \ddots & \\ & & & & \sigma_N^2/\sigma^2 \end{bmatrix}$$

$$= \sigma^2 \, \Omega_{N \times N}$$

$$\boxed{\begin{aligned}\hat{\beta}_{GLS} &= (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y) \\ V(\hat{\beta}_{GLS}) &= \sigma^2(X'\Omega^{-1}X)\end{aligned}}$$

$\hat{\beta}_{OLS} = (X'X)^{-1}X'y$

$V(\hat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}$

$\downarrow$

$\hat{\beta}_{GLS}$ is BLUE ✓ ✓ ✓ ✓

$$V(\hat{\beta}_{GLS}) \leq V(\hat{\beta}_{OLS})$$