# Lecture Note 4: Heteroskedasticity and Dependence

OLS estimator:

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

$$\hat{\beta}_1 = \frac{1}{\sum_i (X_i - \bar{X})^2} \sum_i (Y_i - \bar{Y})(X_i - \bar{X})$$

$$\hat{\beta}_1 = \beta_1 + \underbrace{\frac{1}{\sum_i (X_i - \bar{X})^2} \sum_i (X_i - \bar{X}) U_i}$$

$E[U_i] = 0 \rightarrow \hat{\beta}_1$ unbiased

$\rightarrow 0$ as $N \rightarrow \infty \rightarrow \hat{\beta}_1$ consistent

$$V[\hat{\beta}_1] = \overset{0}{\cancel{V[\beta_1]}} + V\left[\frac{1}{\sum_i (X_i - \bar{X})^2} \sum_i (X_i - \bar{X}) U_i\right]$$

$$V[\hat{\beta}_1] = \left(\frac{1}{\sum_i (X_i - \bar{X})^2}\right)^2 \left[\sum_i (X_i - \bar{X})^2 V[U_i] + \sum_i \sum_{j \neq i} (X_i - \bar{X})(X_j - \bar{X}) \operatorname{cov}(U_i, U_j)\right]$$

Classical model: G-M assumptions

① $E[U_i] = 0$

② $V[U_i] = \sigma^2$

③ $\mathrm{cov}(U_i, U_j) = 0 \quad i \neq j$

$$V[\hat{\beta}_1] = \left(\frac{1}{\sum_i (X_i - \bar{X})^2}\right)^2 \left[\sum_i (X_i - \bar{X})^2 \underbrace{V[U_i]}_{\sigma^2} + \cancel{\sum_i \sum_{j \neq i} (X_i - \bar{X})(X_j - \bar{X}) \mathrm{cov}(U_i, U_j)}\right]$$

→ simple formula for SE

→ in large samples, form $t = \dfrac{\hat{\beta}_1 - \beta_1^0}{SE}$, compare with critical values from $N(0,1)$

## Normal linear model

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

$$U_i \sim N(0, \sigma^2)$$

$\rightarrow$ then $\hat{\beta}_1$ has $t(N-2)$

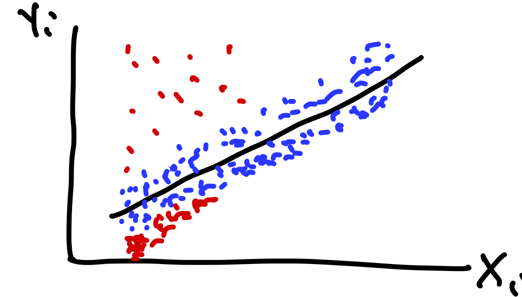
## Random X's

$$E[U_i] = 0 \rightarrow E[U_i \mid X_1, X_2 \cdots X_N] = 0$$

$$E[\hat{\beta}_1] = \beta_1 \rightarrow E[\hat{\beta}_1 \mid X_1, X_2, \cdots, X_N] = \beta_1$$

# Heteroskedasticity

1. $E[U_i | X_i] = 0$
2. $(X_i, Y_i)$ iid
3. outliers unlikely




$$V[\hat{\beta}] = \left(\frac{1}{\sum_i (X_i - \bar{X})^2}\right)^2 \left[\sum_i (X_i - \bar{X})^2 V[U_i] + \cancel{\sum_i \sum_{j \neq i} (X_i - \bar{X})(X_j - \bar{X}) \operatorname{cov}(U_i, U_j)}\right]$$

# Dependence

① clustered sample design

② group-level treatment

$U_i \perp U_j$ across clusters

not within them

$$\hat{U}_i \hat{U}_j$$

$$V[\hat{\beta}_1] = \left(\frac{1}{\sum_i (X_i - \bar{X})^2}\right)^2 \left[\sum_i (X_i - \bar{X})^2 \underbrace{V[U_i]}_{\text{hetero}} + \sum_i \sum_{\cancel{j \neq i}} (X_i - \bar{X})(X_j - \bar{X}) \operatorname{cov}(U_i, U_j)\right]$$

$j \in i$'s cluster

"cluster-robust SE"

feols( )

~~vcov = 'hetero'~~

$\boxed{\text{vcov} = \sim \text{clustvar}}$

WLS: $\hat{\beta}_1^{WLS} = \dfrac{\sum_i w_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i w_i (X_i - \bar{X})^2}$

→ set $w_i$ to account for heteroskedasticity: $w_i = \dfrac{1}{V[U_i]}$

→ known heteroskedasticity

→ grouped data

individual: $Y_{ig} = \beta_0 + \beta_1 X_{ig} + U_{ig}$ $\quad V[U_{ig}] = \sigma^2$

group: $\bar{Y}_g = \beta_0 + \beta_1 \bar{X}_g + \bar{U}_g$ $\quad V[\bar{U}_g] = \dfrac{\sigma^2}{N_g}$

→ WLS: $w_g = \dfrac{1}{V[\bar{U}_g]} = \dfrac{1}{\sigma^2/N_g} = \dfrac{N_g}{\sigma^2}$ → can just set $w_g = N_g$