Lecture 4

OLS estimator:

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

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

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

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

$\rightarrow 0 \rightarrow \hat{\beta}_1$ consistent

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

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

$\operatorname{cov}(aX+bY) = a^2 V[X] + b^2 V[Y] + 2ab\operatorname{cov}(X,Y)$

Classical model

→ GM assumptions: ① $E[U_i] = 0$

② $V[U_i] = \sigma^2$

③ $cov(U_i, U_j) = 0 \quad i \neq j$

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

→ simple formula for SE

→ in large samples, compare $t = \frac{\hat{\beta}_1 - \beta_1^0}{SE}$ w/ 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 a $t(N-2)$

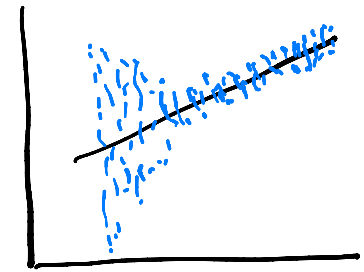
Random X's

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

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

Heteroskedasticity

1. $E[u_i \mid x_i] = 0$
2. (x_i, y_i) iid
3. outliers unlikely



u_i^2

$$V[\hat{\beta}_1] = \left(\frac{1}{\sum_i (x_i - \bar{x})^2}\right)^2 \left[\sum_i (x_i - \bar{x})^2 V[u_i] + \cancel{\sum_i \sum_{j \neq i} (x_i - \bar{x})(x_j - \bar{x}) \operatorname{cov}(u_i, u_j)}\right]$$

vcov = 'hetero'

Dependence

1. clustered sample design
2. group-level treatment

$U_i \perp U_j$ across clusters

<u>not</u> within clusters

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

vcov = ~clustvar

"cluster-robust SE"

$$\bar{Y}_k = \beta_0 + \beta_1 \bar{X}_k + \bar{U}_k$$

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

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

→ WLS: $w_g = \dfrac{N_g}{\sigma^2}$ → can just set $\boxed{w_g = N_g}$