1. Motivation

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

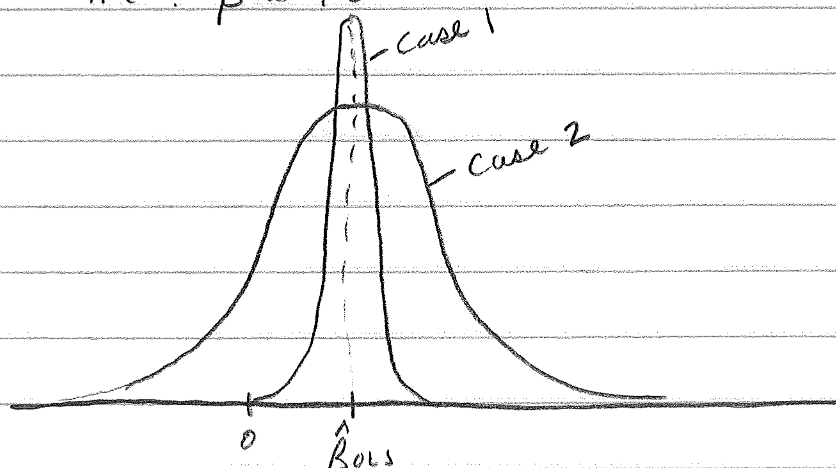
$$\hat{Var}(\hat{\beta}_{OLS}) = \hat{\sigma}^2 (X'X)^{-1} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-K} (X'X)^{-1}$$

$$\hat{SE} = \sqrt{\hat{Var}(\hat{\beta}_{OLS})}$$

Suppose testing:

$H_0 : \hat{\beta}_{OLS} = 0$

$H_a : \hat{\beta}_{OLS} \neq 0$



1. Obtain estimate of β.
2. Is this estimate different from zero?
   a. <u>Case 1</u>. Yes. Standard errors are pretty tight. Reject null in favor of alternate.
   b. <u>Case 2</u>. No. Standard errors are large. Cannot reject at "high enough" confidence level that $\hat{\beta}_{OLS}$ different from zero.

Example: HIV testing of a batch of blood.

→ If underestimating the variance of $\hat{\beta}_{OLS}$, might <u>falsely</u> lead to too tight of standard errors, rejecting H_0 that blood HIV-infected in favor of alternate, not-infected. We want to make sure we don't falsely reject H_0 just because we underestimated $\hat{Var}(\hat{\beta}_{OLS})$.

What can lead to underestimating $\hat{Var}(\hat{\beta}_{OLS})$?

Violation of first assumption of OLS:

(1) $\varepsilon_i \sim iid(0, \sigma^2) \Rightarrow E(\varepsilon)=0,\ Var(\varepsilon)=\sigma^2 \cdot I_{N\times N}$.

a. ε_i not iid

b. $Var(\varepsilon) \neq \sigma^2 \cdot I_{N\times N}$

2. Calculating standard errors under homoskedasticity.

$$Var(\hat{\varepsilon}) = E(\hat{\varepsilon}^2) = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-K} = \frac{\hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \ldots + \hat{\varepsilon}_n^2}{n-K} = \hat{\sigma}^2$$

$\Rightarrow$ Covariance - Variance matrix of ε:

$$\begin{bmatrix} \hat{\sigma}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{\sigma}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \hat{\sigma}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{\sigma}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat{\sigma}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \hat{\sigma}^2 \end{bmatrix}$$

Under homoskedasticity:

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

$$\hat{Var}(\hat{\beta}_{OLS}) = \frac{\hat{\sigma}_2}{Var(\tilde{X}_i)}$$

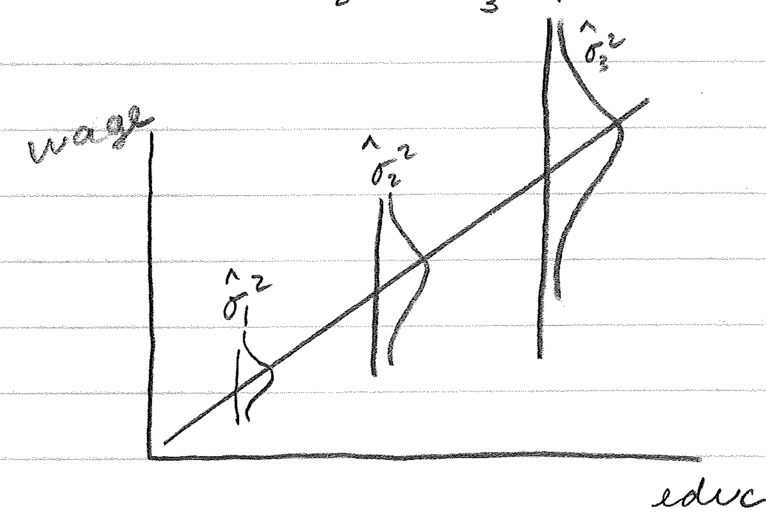
residual of X_i on X_k's for $i \neq k$

3. What is heteroskedasticity?

Covariance - variance matrix of ε:

$$\begin{bmatrix} \hat\sigma_1^2 & 0 & \cdots & & 0 \\ 0 & \hat\sigma_2^2 & & & \vdots \\ \vdots & & \hat\sigma_3^2 & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \hat\sigma_n^2 \end{bmatrix}$$

Here $\hat\sigma_1^2 \neq \hat\sigma_2^2 \neq \hat\sigma_3^2 \neq \ldots \neq \hat\sigma_n^2$



If we just assume $\hat\sigma^2 = \hat\sigma_1^2 = \hat\sigma_2^2 = \ldots = \hat\sigma_n^2$, then $\hat{Var}(\hat\beta_{OLS})$ is usually too small. So correct with the following:

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

$$\hat{\Sigma} = diag[\hat{\varepsilon}_1^2, \hat{\varepsilon}_2^2, \hat{\varepsilon}_3^2, \ldots \hat{\varepsilon}_n^2]$$

$$\boxed{\begin{aligned} &\rightsquigarrow \text{when } \hat{\Sigma} = diag[\hat{\sigma}^2, \hat{\sigma}^2, \hat{\sigma}^2, \ldots \hat{\sigma}^2] \\ &= \cancel{(X'X)^{-1}}\cancel{(X'X)}\hat{\sigma}^2 (X'X)^{-1} \\ &= \hat{\sigma}^2 (X'X)^{-1} \end{aligned}}$$

How to test for heteroskedasticity:

<u>White test:</u>

- regress $\hat{\varepsilon}_i^2$ on X's, squares and cross products
- obtain R^2

Under Ho: No heteroskedasticity

$$\left\{ \; nR^2 \xrightarrow{d} \chi^2(q) \qquad q = K - 1 \text{(constant)} \right.$$

# of regressors

This is known as a LM test.

- if $nR^2 > 5\%$ critical value, then reject Ho.

In stata, "estat imtest, white" after estimation.

Breusch-Pagan:

- similar to White test, except depends on knowing variables causing hetero.

In stata, "estat hettest [variables], iid" after estimation.

variables you think cause hetero, or just use $\hat{y}$ or x's.

↳ White test is the better test because does not assume knowledge of what's causing heteroskedasticity.

Plot and look at the data!

In general, in stata:

(1) reg y x
(2) reg y x, robust

choose the one that maximizes the std errors on your variables of interest.

↳ when heteroskedasticity not present, correcting for it can actually underestimate std errors!

Heteroskedasticity

```
cd "C:\Users\lfeler1\Documents\Applied Econometrics Course\Notes\Weeks 1-3"
clear
clear matrix
set seed 1000
use restricted92
sample 100, count


gen wage=exp(lnw)

*********************************

reg wage ed exp exp2
predict e_hat, resid

mkmat e_hat, matrix(E)
matrix VCV=E*E'

svmat VCV


matrix V=[vecdiag(VCV)]'

svmat V

sum V1
local Vmean=r(mean)

reg ed exp exp2
predict ed_resid, resid
sum ed_resid

display sqrt((`Vmean'*100/96)/((r(sd)^2)*100))

gen constant=1
mkmat ed exp exp2 constant, matrix(X)

matrix VarB=((E'*E)/96)*inv(X'*X)
matrix VB=[vecdiag(VarB)]'
svmat VB
gen SE=sqrt(VB1)


************************************
**Manually calculate robust standard errors**

reg wage ed exp exp2, robust
reg wage ed exp exp2,

matrix sigma=diag(vecdiag(VCV))

matrix VarBr=inv(X'*X)*(X'*sigma*X)*inv(X'*X)

matrix Vr=[vecdiag(VarBr)]'
svmat Vr
gen SEr=sqrt(Vr1)



************************************
**What does heteroskedasticity look like**

twoway (scatter V1 ed) (lfit V1 ed)
```

```
twoway (scatter V1 exp) (lfit V1 exp)

************************************
**How to test for heteroskedasticity**

*White*

reg wage ed exp exp2
estat imtest, white

gen e_hat2=e_hat^2

gen ed2=ed^2
gen exp3=exp^3
gen exp4=exp^4
gen edXexp=ed*exp
gen edXexp2=ed*exp2

reg e_hat2 ed exp exp2 ed2 exp4 edXexp exp3 edXexp2

display e(N)*e(r2)

**Look in a Chi-squared table for this value with 8 degrees of freedom**
display 1-chi2(8,6.6887103)



*Breusch-Pagan*

reg wage ed exp exp2
predict wage_hat
estat hettest, iid

reg e_hat2 wage_hat

display e(N)*e(r2)

**Look in a Chi-squared table for this value with 8 degrees of freedom**
display 1-chi2(1,2.8132943)


        **but if we think a particular variable is causing heteroskedasticity, we
can do the
        **Breusch-Pagan test just on that

reg wage ed exp exp2
estat hettest ed, iid

reg e_hat2 ed

display e(N)*e(r2)

**Look in a Chi-squared table for this value with 8 degrees of freedom**
display 1-chi2(1,3.6802517)
```

3. Correcting SEs for clustering

$E(\varepsilon_i \cdot \varepsilon_j) \neq 0 \Rightarrow$ clustering, random group effects
$\Rightarrow$ serial correlation (time series)

$Var(\varepsilon) = \Sigma$, off-diagonal elements $\neq 0$.

If $E(\varepsilon_i \cdot \varepsilon_j) > 0$ (positive correlation between errors), then w/o correction for clustering, $\hat{Var}(\hat{\beta}_{OLS})$ is biased down (too small).

If $E(\varepsilon_i \cdot \varepsilon_j) < 0$ (negative correlation between errors), then w/o correction for clustering, $\hat{Var}(\hat{\beta}_{OLS})$ is biased up (too big).

Example:

$y_{is} = \beta X_s + \varepsilon_{is}$

$$\varepsilon_{is} = a_s + u_{is}$$

$u_{is} \sim iid(0, \sigma_u^2)$

random school effect $\sim (0, \sigma_s^2)$

- individuals in same schools have similar unobservables, shocks

$$E(a_s \cdot u_{is}) = 0$$

$$E(\varepsilon_{is} \cdot \varepsilon_{js}) = \sigma_s^2 > 0$$

$$E(\varepsilon\varepsilon') = \Sigma = \begin{bmatrix} \begin{matrix} \ddots & \sigma^2_{s1} \\ \sigma^2_{s1} & \ddots \end{matrix} & 0 & 0 \\ 0 & \begin{matrix} \ddots & \sigma^2_{s2} \\ \sigma^2_{s2} & \ddots \end{matrix} & 0 \\ 0 & 0 & \begin{matrix} \ddots & \sigma^2_{sn} \\ \sigma^2_{sn} & \ddots \end{matrix} \end{bmatrix} \equiv \text{Block diagonal with S blocks}$$

$$\hat{Var}(\hat{\beta}_{OLS,c}) = (X'X)^{-1}(X'\hat{\Sigma}X)(X'X)^{-1}$$

1. estimate $\hat{\sigma}_s^2$ for $s = 1, \ldots, n$ and plug into $\hat{\Sigma}$. **OR**
2. in stata:

   reg y x, cluster(school)

   or
3. $\bar{y}_s = \beta X_s + \bar{\varepsilon}_s$ and weight by N_s

   $\bar{\varepsilon}_s = \bar{u}_s \sim iid(0, \sigma^2_{\bar{u}})$

   ↳ OLS with clustering is more efficient, but not always possible. # of clusters must be > 42!

4. Bootstrap standard errors (for this, number of clusters can be less than 42).

What is bootstrapping?

↳ Suppose 400 random samples of an estimator $\hat{\beta}$ were available from the population. Then to get the "standard error" of $\hat{\beta}$, we could simply calculate the standard deviation of the 400 $\hat{\beta}$'s.

→ Bootstrapping will draw a random sample (or clustered random sample) from your <u>sample</u> 400 times, calculate 400 $\hat{y} = X'\hat{\beta}$, and then take the standard deviation of these 400 $\hat{\beta}$'s and report that as the "standard error."

↳ This is computationally intensive!

Let $\hat{\beta}_1, \ldots, \hat{\beta}_{400}$ denote the 400 estimates $\hat{\beta}$:
The bootstrap estimate of the variance of $\hat{\beta}$ is:

$$\hat{Var}_{boot}(\hat{\beta}) = \frac{1}{400-1} \sum_{b=1}^{400} (\hat{\beta}_b - \bar{\hat{\beta}})^2$$

$$\bar{\hat{\beta}} = \frac{1}{400} \sum_{b=1}^{400} \hat{\beta}_b . \rightarrow \text{Just the average of all the 400 } \hat{\beta}\text{'s.}$$

Why 400? Apparently, if you increase beyond 400, you don't decrease the $\hat{Var}_{boot}(\hat{\beta})$ by all that much, but you increase computation time.

Clustering and Bootstrap

```
clear
use prosp

reg nmsc wc

***Clustering at the school level: note the number of clusters***
sort schoolid
loneway nmsc schoolid
reg nmsc wc, cluster(schoolid)


***Just obtain school means for y and x; now no worries about clustered SEs***
preserve
collapse (mean) nmsc wc, by(schoolid)

reg nmsc wc

restore


***Bootstrap standard errors, with resampling iid across clusters but not iid within
clusters***
reg nmsc wc, vce(boot, cluster(schoolid) reps(400) seed(1000))
```