

# Standard Errors

---

**Venki**

PhD student

Public Policy and Political Economy, School of EPPS

---

# Start recording!

---

**Note to self**

## Outline

---

**1**

*Deriving the variance-covariance matrix of OLS regression*

**2**

*Homoskedasticity & Spherical errors*

**3**

*Heteroskedasticity robust errors*

**4**

*Cluster robust errors*

**5**

*Takeaways*

**6**

*Worksheet answers discussion*

## 1. Deriving the variance-covariance matrix of OLS regression

# Causal effect is a regression coefficient

*Remember this slide?*

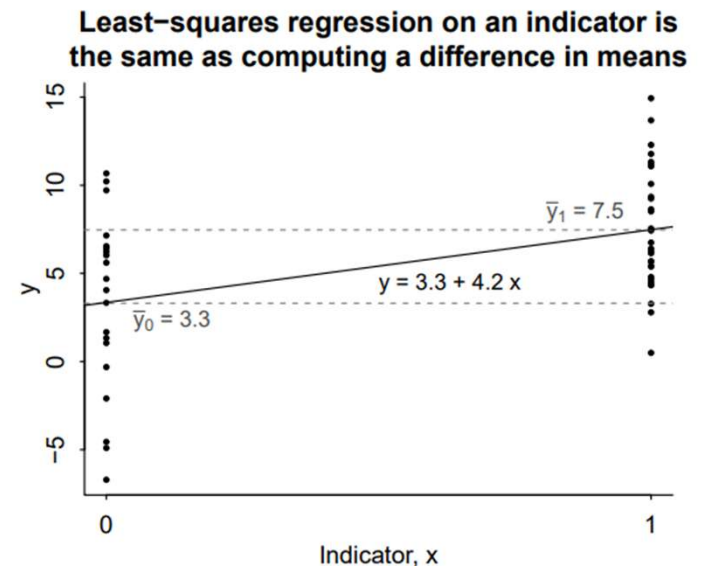
### Estimator (Regression with Experiments)

The ATE can be expressed as a regression coefficient:

$$\begin{aligned} Y_i &= D_i Y_{1i} + (1 - D_i) Y_{0i} \\ &= Y_{0i} + (Y_{1i} - Y_{0i}) D_i \\ &= \underbrace{\bar{Y}_0}_{\alpha} + \underbrace{(\bar{Y}_1 - \bar{Y}_0)}_{\tau_{Reg}} D_i + \underbrace{\{(Y_{i0} - \bar{Y}_0) + D_i \cdot [(Y_{i1} - \bar{Y}_1) - (Y_{i0} - \bar{Y}_0)]\}}_{\varepsilon_i} \\ &= \alpha + \tau_{Reg} D_i + \varepsilon_i \end{aligned}$$

*Now... how to get the SE of the regression coefficient?*

Only then we can get SE of the causal effect!



Gelman, A., Hill, J., & Vehtari, A.  
(2020). *Regression and other stories*.  
Cambridge University Press.

## 1. Deriving the variance-covariance matrix of OLS regression

---

# How to get the SE of betas?

*To get the standard errors of the regression coefficient, we need...*

*Why? What is this?  
Let us derive it.*

All a"board"!

$$\begin{bmatrix} E[\epsilon_0^2|X] & E[\epsilon_0\epsilon_1|X] & \cdots & E[\epsilon_0\epsilon_n|X] \\ E[\epsilon_1\epsilon_0|X] & E[\epsilon_1^2|X] & \cdots & E[\epsilon_1\epsilon_n|X] \\ \vdots & \vdots & \ddots & \vdots \\ E[\epsilon_n\epsilon_0|X] & E[\epsilon_n\epsilon_1|X] & \cdots & E[\epsilon_n^2|X] \end{bmatrix}$$

## 1. Deriving the variance-covariance matrix of OLS regression

---

# Bottomline

*From the math, you see that to get the standard errors of the regression coefficient, we need...*

$$E[\epsilon^T \epsilon | X] = \begin{bmatrix} E[\epsilon_0^2 | X] & E[\epsilon_0 \epsilon_1 | X] & \cdots & E[\epsilon_0 \epsilon_n | X] \\ E[\epsilon_1 \epsilon_0 | X] & E[\epsilon_1^2 | X] & \cdots & E[\epsilon_1 \epsilon_n | X] \\ \vdots & \vdots & \ddots & \vdots \\ E[\epsilon_n \epsilon_0 | X] & E[\epsilon_n \epsilon_1 | X] & \cdots & E[\epsilon_n^2 | X] \end{bmatrix}$$

*But this is filled with population parameters – the errors!*

*All we can do now is make assumptions on the structure of this matrix, to enable estimation.*

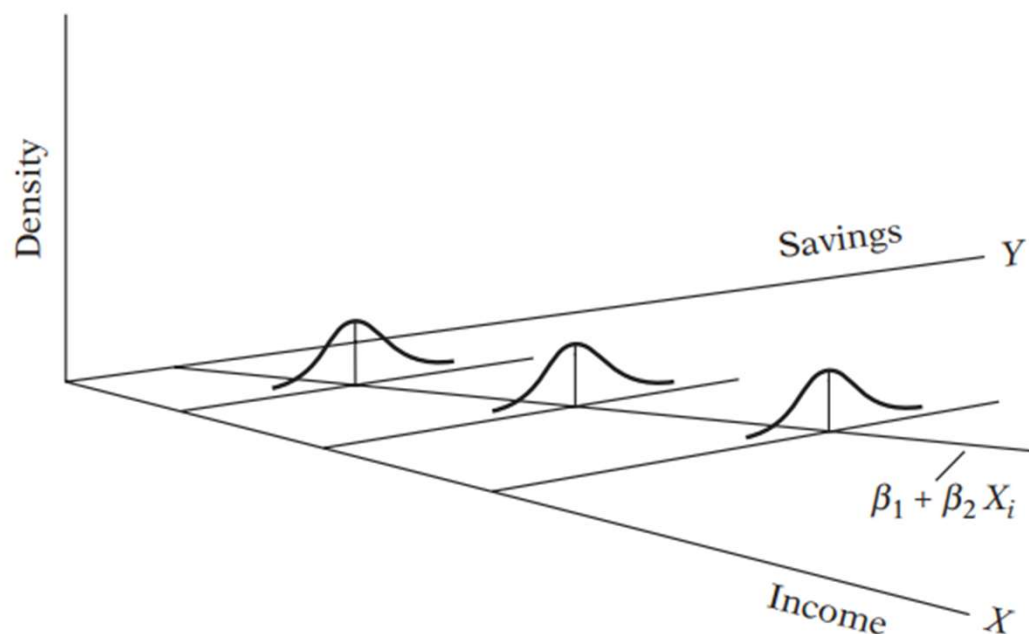
*Remember...*

All of this has NOTHING to do with the bias. OLS estimator is still unbiased.

U in the “BLUE” is still fine...

## 2. Homoskedasticity & Spherical errors

# Simplest Assumption: Errors are homoskedastic



### *“Homoskedasticity”*

“Homo”: Same

“Skedasticity”: Like ‘skedaddle’

So, it means spread out similarly

*Errors in all observations have the same variance.*

And no covariance.

## 2. Homoskedasticity & Spherical errors

---

# Estimating Spherical errors

*Homoskedasticity assumption leads to what are called “spherical errors”*

$$E[\epsilon\epsilon^\top | X] = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}$$
$$\implies V[\hat{\beta} | X] = \sigma^2 (X^\top X)^{-1}$$

*But there is still a population parameter! Luckily, we have an estimator\* for it...*

$$\hat{\sigma}^2 = \frac{\mathbf{e}^\top \mathbf{e}}{n - p} \implies \widehat{V[\hat{\beta} | X]} = \frac{\mathbf{e}^\top \mathbf{e}}{n - p} (X^\top X)^{-1}$$

*This is what `lm` (by default) gives in R!*

So, `lm` (by default) makes the homoskedasticity assumption!

\* - I urge you to lookup the properties of this estimator. Is it unbiased?

## 2. Homoskedasticity & Spherical errors

# How to justify making the homoskedasticity assumption?

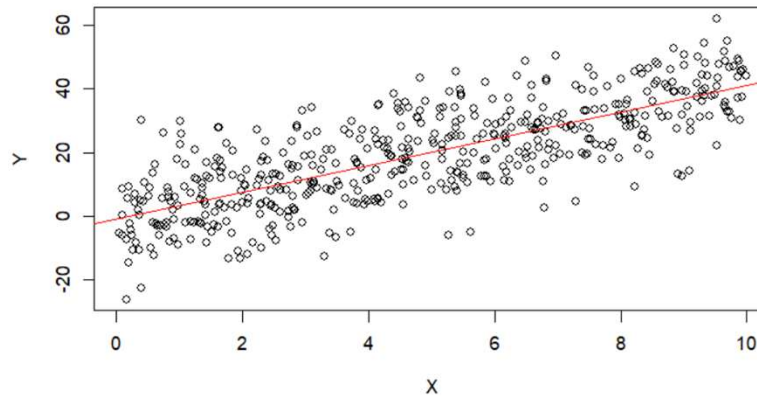
*YOU CAN'T!*

There are supposed tests and you can eyeball a residual plot, but you can never be guaranteed.

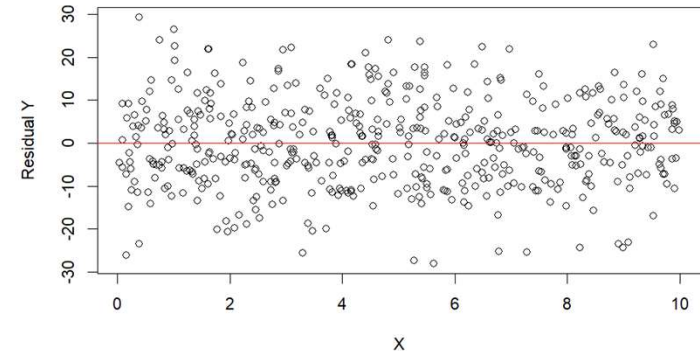
So just don't make such highly limiting and strong assumptions!

*Don't report the SE  
lm (by default) gives  
you!*

Plot of Observations with Homoskedastic errors/residuals

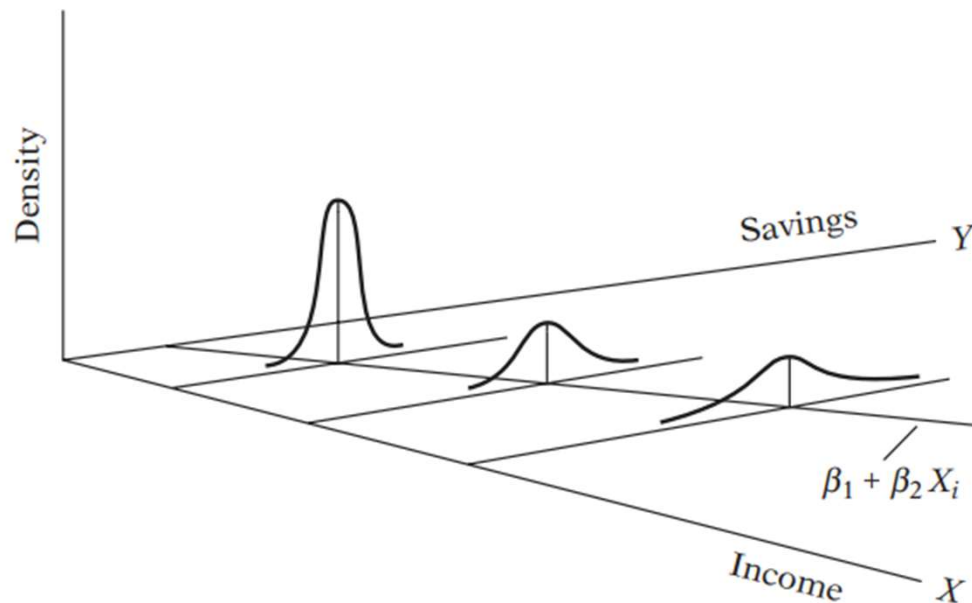


Residual Plot of Fitted Model with Homoskedastic Errors



### 3. Heteroskedasticity robust errors

# Heteroskedasticity



*We can happily allow for different error variance for each observation.*

In fact, you have already done this!

$$SE_{DiM} = \sqrt{\frac{V(X_t)}{N_t} + \frac{V(X_c)}{N_c}}$$

### 3. Heteroskedasticity robust errors

## Estimating Heteroskedasticity robust errors

*We now need heteroskedasticity robust errors.*

$$E[\epsilon\epsilon^\top | X] = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

$$\Rightarrow V[\hat{\beta}|X] = (X^\top X)^{-1} \frac{1}{n} \sum_{i=1}^n \sigma_i^2 \mathbf{x}_i \mathbf{x}_i^\top (X^\top X)^{-1}$$

*We have multiple estimators for this accounting for different finite sample adjustments, leverage etc.*

HC0:  $\widehat{V[\hat{\beta}|X]}_{HC0} = (X^\top X)^{-1} X^\top \text{diag}[e_i^2] X (X^\top X)^{-1}$

HC1:  $\widehat{V[\hat{\beta}|X]}_{HC1} = \frac{n}{n-p} (X^\top X)^{-1} X^\top E[\text{diag}[e_i^2]] X (X^\top X)^{-1}$

HC2:  $\widehat{V[\hat{\beta}|X]}_{HC2} = (X^\top X)^{-1} X^\top \text{diag}\left[\frac{e_i^2}{1-h_{ii}}\right] X (X^\top X)^{-1}$

HC3:  $\widehat{V[\hat{\beta}|X]}_{HC3} = (X^\top X)^{-1} X^\top \text{diag}\left[\frac{e_i^2}{1-h_{ii}^2}\right] X (X^\top X)^{-1}$

### 3. Heteroskedasticity robust errors

---

## Bottomline - Report HC2

*Understand the ideas without worrying too much about the magic with this one\*.*

*But worry about the code!*

*Generally, you are required to report HC2 errors.*

Stata reports HC1 by default.

In R,

**DON'T** use `lm` with default parameters, because it reports the homoscedastic errors!

Instead get HC2 errors like so:

```
library(estimatr)
hc2_lm = lm_robust(Y ~ X - 1, se_type = "HC2")
```

\*This is about the previous slide only. All the other slides before it, please practice the math at least once in your life!

#### 4. Cluster robust errors

---

## Similar units within a cross section

| City      | Country | Other covariates |
|-----------|---------|------------------|
| Chennai   | India   | ...              |
| Bengaluru | India   | ...              |
| Dallas    | USA     | ...              |
| Austin    | USA     | ...              |

*One can imagine that observations within the same country are similar to each other but dissimilar to other countries. In this sense there are clusters in the data.*

*So, what would the errors look like now?*

#### 4. Cluster robust errors

Clustered errors have block diagonal structure

$$E[\epsilon\epsilon^\top|X] = \begin{bmatrix} \sigma_{(1,1)1}^2 & \cdots & \sigma_{(1,n_1)1}^2 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{(n_1,1)1}^2 & \cdots & \sigma_{(n_1,n_1)1}^2 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_{(1,1)2}^2 & \cdots & \sigma_{(1,n_2)2}^2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_{(n_2,1)2}^2 & \cdots & \sigma_{(n_2,n_2)2}^2 \\ & & & & \ddots & \\ & & & & & \sigma_{(1,1)m}^2 & \cdots & \sigma_{(1,n_m)m}^2 \\ & & & & & \vdots & \ddots & \vdots \\ & & & & & \sigma_{(n_m,n_m)m}^2 & \cdots & \sigma_{(n_m,n_m)m}^2 \end{bmatrix}$$

$$V[\hat{\beta}|X] = (X^\top X)^{-1} \sum_{g=1}^m \mathbf{x}_g^\top \epsilon_g \epsilon_g^\top \mathbf{x}_g (X^\top X)^{-1}$$

#### 4. Cluster robust errors

---

## Estimating Cluster robust errors

### *Same story*

Don't worry too much about the math here. But worry about the code.

### *The math*

$$\widehat{V[\hat{\beta}|X]}_{CR_{fancy}} = \frac{m}{m-1} \frac{n-1}{n-p} (X^\top X)^{-1} \sum_{g=1}^m \mathbf{x}_g^\top \mathbf{e}_g \mathbf{e}_g \mathbf{x}_g (X^\top X)^{-1}$$

### *The code*

Stata reports this.

In R,

```
library(estimatr)
cr_lm = lm_robust(my_dv ~ X1 + X2 + X3,
                  clusters = countries,
                  se_type = "CR2")
```

## 5. Takeaways

---

# Remember these salient points

*Don't report the default  $1m$  errors!*

Don't make the homoskedasticity assumption

Don't rely on those homo/heteroskedasticity tests.

FYI these are called Breusch-Pagan tests. See the Gujarati book for an excellent explanation of its limitations.

*Don't try all the different errors and then pick the best! That amounts to p-hacking!*

Just use HC2. Or pre-register what you are going to report.

*This is all about the SE and NOT bias.*

OLS is still unbiased!

Choice of errors only affects variance of the regression coefficient not the estimate itself,

*Clustering errors requires domain knowledge.*

## Outline

---

# Proceeding to worksheet 2...

**1**

*Deriving the variance-covariance matrix of OLS regression* ✓

**2**

*Homoskedasticity & Spherical errors* ✓

**3**

*Heteroskedasticity robust errors* ✓

**4**

*Cluster robust errors* ✓

**5**

*Takeaways* ✓

**6**

*Worksheet answers discussion*

## 6. Worksheet answers discussion

---

*Let us now take a step back and discuss the answers for worksheet 2*

*All a”board”!*

As I had announced on elearning we awarded full points for everyone on the power calculation question (2.2).

Here are the questions that more than half the submissions got wrong: 1a, 1c, 3b, 2.2

## 6. Worksheet answers discussion

---

Good job on the worksheet!

*Class average has improved even without 2.2 adjustment! Keep it up!*



Everything all at once

---

# Summary

## Deriving the variance-covariance matrix of OLS regression

- Practice this math.
- Take a pic and share in Teams group for bragging rights!

## Homoskedasticity & Spherical errors

- Highly limiting assumption.
- Don't use default Se from lm!

## Heteroskedasticity robust errors

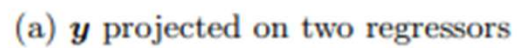
- Generally, report HC2 errors

## Cluster robust errors

- Understand your domain to figure out how to cluster errors

## Worksheet

- Pay attention to power when you do the datacamp (if you haven't already!)



See *Econometric Theory and Methods* by Davidson and MacKinnon. ([Free online book!](#))  
That whole book is written with Matrix notation and FWL is introduced at the top of that book.

See you later in class!