TD 1:
OLS Regressions
ECO 567A

Geoffrey Barrows[1]

[1]CREST, CNRS, Ecole polytechnique

Jan 10, 2025

Outline

1. What is the OLS estimator? What does it let you say? When is it valid?
2. Derive OLS estimator
3. Compute the variance of OLS estimator
4. Construct confidence intervals
5. Read regression tables

Dell Jones Olken (2012)

1. What is the relationship between temperature and GDP growth?
2. In particular, is the relationship stronger for poor countries?

Example of a Regression Table

$$g_{it} = \theta_i + \theta_t + \beta_T * T_{it} + \gamma_T * T_{it} * 1(poor) + \beta_P * P_{it} + \gamma_p * P_{it} * 1(poor) + \epsilon_{it}$$

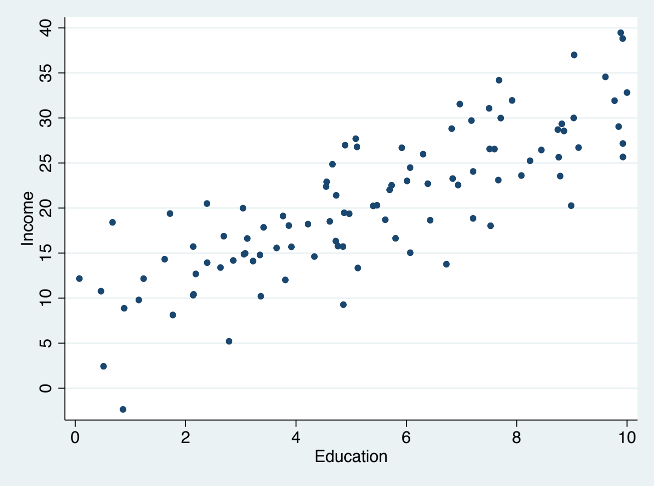
TABLE 2—MAIN PANEL RESULTS

Dependent variable is the annual growth rate	(1)	(2)	(3)	(4)	(5)
Temperature	−0.325	0.261	0.262	0.172	0.561*
	(0.285)	(0.312)	(0.311)	(0.294)	(0.319)
Temperature interacted with...					
Poor country dummy		−1.655***	−1.610***	−1.645***	−1.806***
		(0.485)	(0.485)	(0.483)	(0.456)
Hot country dummy				0.237	
				(0.568)	
Agricultural country dummy					−0.371
					(0.409)
Precipitation			−0.083*	−0.228***	−0.105**
			(0.050)	(0.074)	(0.053)
Precipitation interacted with...					
Poor country dummy			0.153*	0.160**	0.145*
			(0.078)	(0.075)	(0.087)
Hot country dummy				0.185**	
				(0.078)	
Agricultural country dummy					0.010
					(0.085)
Observations	4,924	4,924	4,924	4,924	4,577
Within R^2	0.00	0.00	0.00	0.01	0.01
R^2	0.22	0.22	0.22	0.22	0.24
Temperature effect in poor countries		−1.394***	−1.347***	−1.473***	−1.245***
		(0.408)	(0.408)	(0.440)	(0.463)
Precipitation effect in poor countries			0.069	−0.0677	0.0401
			(0.058)	(0.073)	(0.089)

Dell Jones Olken (2012)

1. 1 degree Celsius increase lowers GDP growth in poor countries by 1.394 percentage points
2. This estimate is statistically significant at the 1% level.

Raw Data



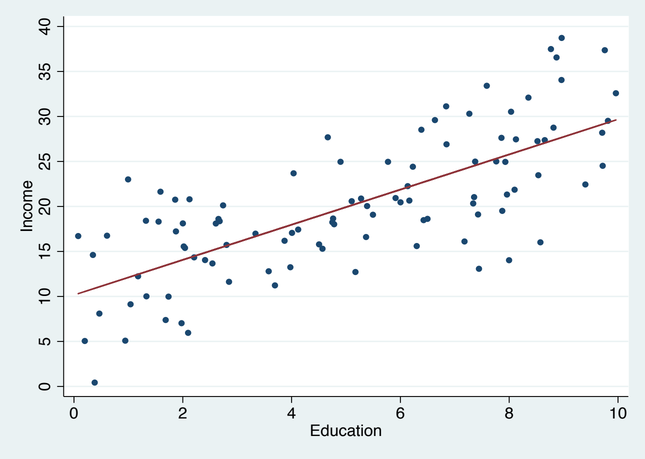
Posit a Model

$$y_i = \alpha + \beta * x_i + \epsilon_i \tag{1}$$

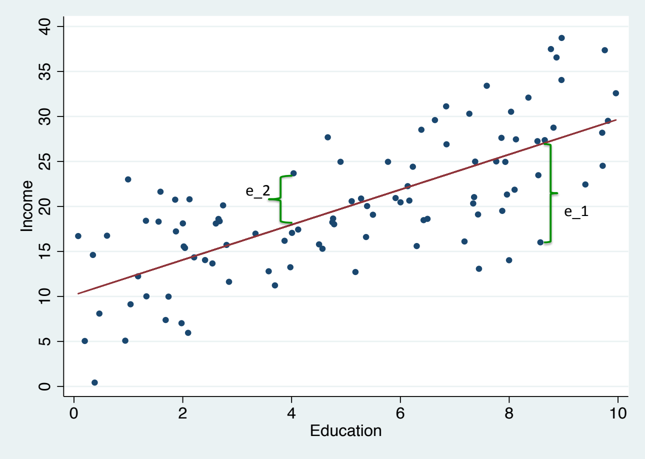
$$y = X\theta + \epsilon \tag{2}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Need to fit a line



How to choose the line?



Estimation Criteria

$$\underset{\theta}{\operatorname{argmin}} \quad (y - X\theta)'(y - X\theta)$$

Review of Linear Algebra

A, B are matrices, a, b are vectors

$$
\begin{aligned}
a'b &= b'a \quad (3)\\
(AB)' &= B'A' \quad (4)\\
(A+B)' &= A' + B' \quad (5)\\
\frac{\partial b'a}{\partial b} &= a \quad (6)\\
\frac{\partial b'Ab}{\partial b} &= 2Ab \quad (7)
\end{aligned}
$$

Solving for OLS Estimator

$$\underset{\theta}{\text{argmin}} \quad (y - X\theta)'(y - X\theta)$$

$$\underset{\theta}{\text{argmin}} \quad [y' - (X\theta)'](y - X\theta)$$

$$\underset{\theta}{\text{argmin}} \quad y'y - y'X\theta - \theta'X'y + \theta'X'X\theta$$

Solving for OLS Estimator

$$\underset{\theta}{\text{argmin}} \quad y'y - y'X\theta - \theta'X'y + \theta'X'X\theta$$

$$\frac{\partial}{\partial\theta}\left(y'X\theta\right) = \frac{\partial}{\partial\theta}\left(\theta'X'y\right) = X'y \quad \text{by (6)}$$

$$\frac{\partial}{\partial\theta}\left(\theta'X'y\right) = X'y \quad \text{by (6)}$$

$$\frac{\partial}{\partial\theta}\left(\theta'X'X\theta\right) = 2X'X\theta \quad \text{by (7)}$$

$$\implies \text{First Order Condition:} - 2X'y + 2X'X\theta = 0$$

$$\implies \hat{\theta} = \frac{X'y}{X'X}$$

The reason to use OLS: Gauss-Markov Theorem

- OLS is the “Best” Linear Unbiased Estimator
  - “Best” means minimum variance

OLS is Unbiased

Re-writing $\hat{\theta}$

$$
\begin{aligned}
\hat{\theta} &= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\theta + \epsilon) \\
&= \theta + (X'X)^{-1}X'\epsilon
\end{aligned}
$$

Taking expectation

$$E[\hat{\theta}] = E[\theta + (X'X)^{-1}X'\epsilon] = \theta$$

Key Assumption

$$E[X'\epsilon] = 0 \quad \text{equivalently} \quad E[\epsilon|X] = 0$$

Example

- model

$$y_i = 10 * x_i + \underbrace{a_i + \epsilon_i}_{u_i}$$

id	x	a	ϵ	u	y
1	1	-1	-1	-2	8
2	1	-3	1	-2	8
3	2	1	-1	0	20
4	2	3	1	4	24

- Notice

$$\begin{aligned} E[X'u] &= (-2-2+0+8)/4 = 1 \\ E[X'\epsilon] &= (-1+1-2+2)/4 = 0 \end{aligned}$$

Example

- model

$$y_i = 10 * x_i + \underbrace{a_i + \epsilon_i}_{u_i}$$

id	x	a	ϵ	u	y
1	1	-1	-1	-2	8
2	1	-3	1	-2	8
3	2	1	-1	0	20
4	2	3	1	4	24

- In Matrix Form

$$\begin{bmatrix} 8 \\ 8 \\ 20 \\ 24 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & -3 \\ 2 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 10 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$$

Run OLS with "a"

- Take X as

$$\begin{bmatrix} 1 & -1 \\ 1 & -3 \\ 2 & 1 \\ 2 & 3 \end{bmatrix}$$

- Applying OLS

$$\begin{bmatrix} \widehat{\theta_1} \\ \widehat{\theta_2} \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & 2 & 2 \\ -1 & -3 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & -3 \\ 2 & 1 \\ 2 & 3 \end{bmatrix} \right)^{-1} * \left(\begin{bmatrix} 1 & 1 & 2 & 2 \\ -1 & -3 & 1 & 3 \end{bmatrix} \begin{bmatrix} 8 \\ 8 \\ 20 \\ 24 \end{bmatrix} \right)$$

$$\begin{bmatrix} \widehat{\theta_1} \\ \widehat{\theta_2} \end{bmatrix} = \frac{1}{184} \begin{bmatrix} 20 & -4 \\ -4 & 10 \end{bmatrix} * \begin{bmatrix} 104 \\ 60 \end{bmatrix} = \begin{bmatrix} 10 \\ 1 \end{bmatrix}$$

Run OLS without "a"

- Take X as

$$\begin{bmatrix}1\\1\\2\\2\end{bmatrix}$$

- Applying OLS

$$\begin{bmatrix}\widehat{\theta_1}\end{bmatrix} = \left(\begin{bmatrix}1 & 1 & 2 & 2\end{bmatrix} \begin{bmatrix}1\\1\\2\\2\end{bmatrix} \right)^{-1} * \left(\begin{bmatrix}1 & 1 & 2 & 2\end{bmatrix} \begin{bmatrix}8\\8\\20\\24\end{bmatrix} \right)$$

$$\begin{bmatrix}\widehat{\theta_1}\end{bmatrix} = \frac{1}{10} * 104 = 10.4$$

Monte Carlo Experiment

- Let true model be $y_i = 10 + 2x_i + \epsilon_i$

- Draw 100 observations

$$\begin{aligned} x_i &\sim U(0,10) \\ \epsilon_i &\sim N(0,25) \\ y_i &= 10 + 2x_i + \epsilon_i \end{aligned}$$

Table: Summary statistics 1

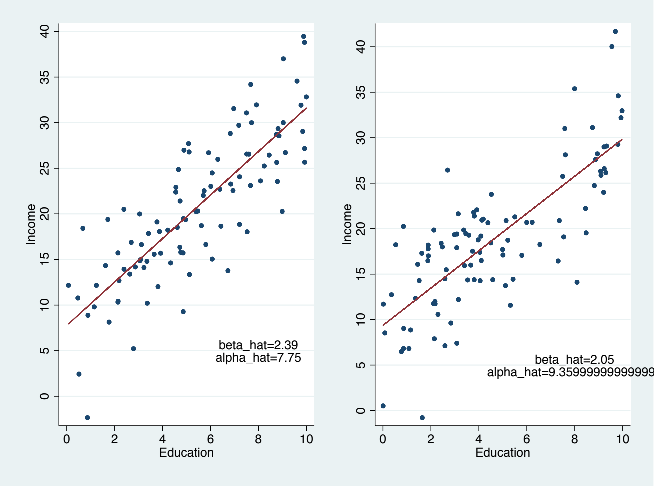
Variable	Mean	Std. Dev.	Min.	Max.	N
y	20.606	7.899	-2.35	39.463	100
x	5.386	2.705	0.074	9.999	100
eps	-0.166	4.671	-14.08	9.693	100

Table: Summary statistics 2

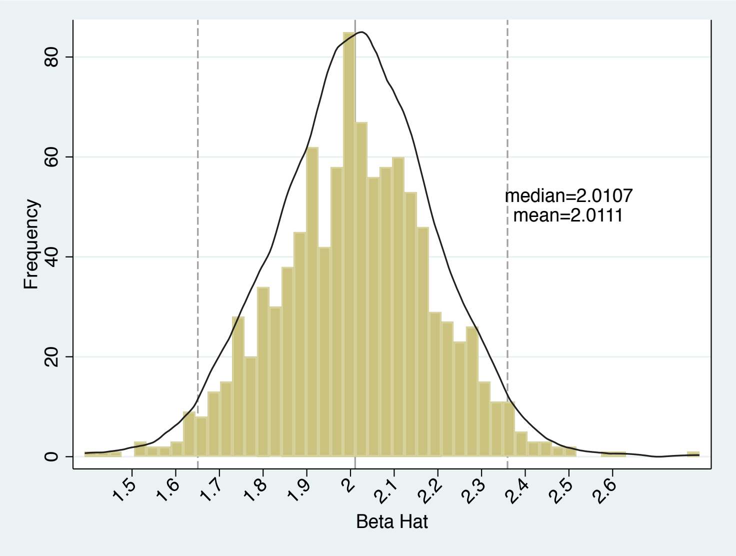
Variable	Mean	Std. Dev.	Min.	Max.	N
y	18.812	7.704	-0.79	41.682	100
x	4.609	2.92	0.008	9.974	100
eps	-0.406	4.851	-14.069	12.292	100

OLS in 2 different realizations of data

Let true model be $y_i = 10 + 2x_i + \epsilon_i$



1000 replications



Hypothesis Testing and Confidence Intervals

- How good is our estimate $\hat{\theta}$?
  Given sampling error, it's never exactly equal to θ

- Given a hypothesis of θ, can we reject that a sample would yield $\hat{\theta}$ with some probability?

Hypothesis Testing and Confidence Intervals

Let $\hat{\mu}$ be a statistical estimate of μ with $\hat{\mu} \sim N[\mu, \frac{\sigma^2}{n}]$.
Then the z-statistic for a population of size n is

$$z = \frac{\hat{\mu} - \mu}{\sqrt{VAR\left(\hat{\mu}\right)}} = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim N[0, 1]$$

And the t-statistic

$$t = \frac{\hat{\mu} - \mu}{\sqrt{\widehat{VAR}\left(\hat{\mu}\right)}} = \frac{\hat{\mu} - \mu}{s/\sqrt{n}} \equiv \frac{\hat{\mu} - \mu}{se(\hat{\mu})} \sim \text{Student t}$$

Null Hypothesis

$$\begin{aligned} H_0 \quad &: \quad \mu = \mu_0 \\ &\Longrightarrow \text{reject if } |t| > |t_{critical}| \end{aligned}$$

95% Confidence Interval

$$\mu_0 \in (\hat{\mu} \pm t_{.025} * se(\hat{\mu}))$$

Interpretation of 95% Confidence Intervals

$$\mu \in (\hat{\mu} \pm t_{.025} * se(\hat{\mu}))$$

- This is the range of values that we would fail to reject the null hypothesis
- The true value might lie outside the range. Can't know for sure.
- 95% of the time, the CI you build should cover the true value
- Implication $\longrightarrow$ 5% of the time, you will mistakenly reject the true value

Variance of OLS Estimator

Taking expectation

$$\begin{aligned} E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] &= E[((X'X)^{-1}X'\epsilon)((X'X)^{-1}X'\epsilon)'] \\ &= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}] \\ &= E[(X'X)^{-1}X'\sigma^2 IX(X'X)^{-1}] \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Key Assumption

$$E[\epsilon\epsilon'] = \sigma^2 I$$

Estimating the Variance of OLS Estimator

In theory

$$E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] = \sigma^2 (X'X)^{-1}$$

Estimate

$$\hat{\sigma}^2 = \frac{e'e}{n-k}$$

Then

$$E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] = \frac{e'e}{n-k} (X'X)^{-1}$$

Hypothesis Testing with OLS Estimator

Assume

$$\epsilon | X \sim N[0, \sigma^2]$$

Then

$$\hat{\theta} \sim N[\theta, \sigma^2 (X'X)^{-1}]$$

Standard Error

$$se(\hat{\theta}) = \sqrt{\frac{e'e}{n-k}(X'X)^{-1}}$$

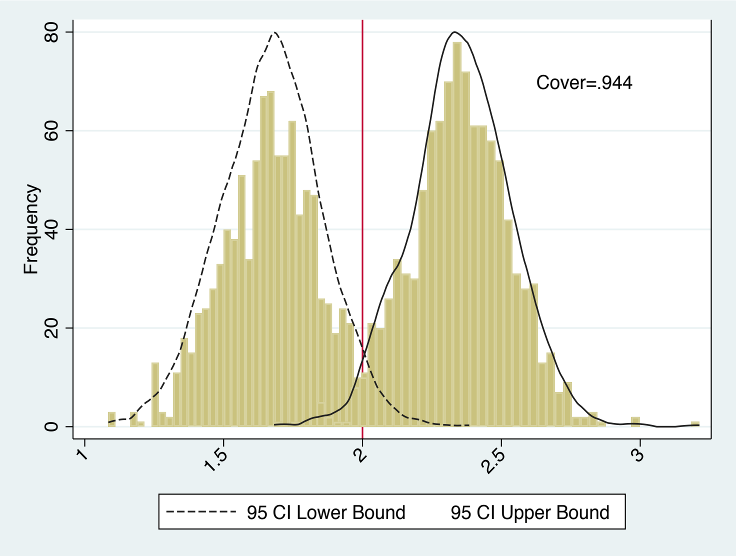
t-statistic

$$\frac{\hat{\theta} - \theta}{se(\hat{\theta})} \sim \text{Student t}$$

95% Confidence Interval

$$\theta \in (\hat{\theta} \pm t_{.025} * se(\hat{\theta}))$$

95% CI for Beta Hat



Example of a Regression Table

$$g_{it} = \theta_i + \theta_t + \beta_T * T_{it} + \gamma_T * T_{it} * 1(poor) + \beta_P * P_{it} + \gamma_p * P_{it} * 1(poor) + \epsilon_{it}$$

TABLE 2—MAIN PANEL RESULTS

Dependent variable is the annual growth rate	(1)	(2)	(3)	(4)	(5)
Temperature	−0.325	0.261	0.262	0.172	0.561*
	(0.285)	(0.312)	(0.311)	(0.294)	(0.319)
Temperature interacted with...					
Poor country dummy		−1.655***	−1.610***	−1.645***	−1.806***
		(0.485)	(0.485)	(0.483)	(0.456)
Hot country dummy				0.237	
				(0.568)	
Agricultural country dummy					−0.371
					(0.409)
Precipitation			−0.083*	−0.228***	−0.105**
			(0.050)	(0.074)	(0.053)
Precipitation interacted with...					
Poor country dummy			0.153*	0.160**	0.145*
			(0.078)	(0.075)	(0.087)
Hot country dummy				0.185**	
				(0.078)	
Agricultural country dummy					0.010
					(0.085)
Observations	4,924	4,924	4,924	4,924	4,577
Within R^2	0.00	0.00	0.00	0.01	0.01
R^2	0.22	0.22	0.22	0.22	0.24
Temperature effect in poor countries		−1.394***	−1.347***	−1.473***	−1.245***
		(0.408)	(0.408)	(0.440)	(0.463)
Precipitation effect in poor countries			0.069	−0.0677	0.0401
			(0.058)	(0.073)	(0.089)

References

- Greene, William H. Econometric analysis. Pearson Education India, 2003
- Angrist, Joshua D., and Jörn-Steffen Pischke. Mostly harmless econometrics: An empiricist's companion. Princeton university press, 2008
- Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. "Temperature shocks and economic growth: Evidence from the last half century." American Economic Journal: Macroeconomics 4.3 (2012): 66-95.