Session 2: OLS (I)

MGT 581 | Introduction to econometrics

Michaël Aklin

PASU Lab | EPFL

Last time...

- Unit Treatment Effect, Average Treatment Effect, Conditional Average Treatment Effect
- Fundamental problem of causal inference
- Baseline bias, differential treatment effect bias

Plan for today

- **Association** statistics between two variables (cov, cor)
- Building the (theoretical) (population) **linear model**
- Estimating the LM from data with **ordinary least squares**
- Readings: Stock and Watson (2011) (ch4), Verbeek (2018) (ch2.1-2.3)
- Also: Cameron and Trivedi (2005) (ch4.1-4.5), Greene (2008) (ch2-3), Casella and Berger (2021) (entire book for statistics)

Background

- **Linear model**: workhorse of econometrics
- Distinguish between **theoretical model** and **estimation technique**
- Linear model is a **theoretical** model...
- ... whose parameters (generally) need to be estimated
- Step 1: build the model
- Step 2: estimate the model

Notation

- $Y \in \mathbb{R}$: outcome, dependent variable, effect, etc.
- Y^0, Y^1: potential outcomes ($\equiv (Y|D = 0), (Y|D = 1)$)
- $D \in \{0, 1\}$: treatment, independent variable, cause, regressor
- $E[\cdot] \equiv \mu$: (population) expectation (unobserved) (1st moment)
  - $E[\cdot] \equiv \sum_{i=1}^{n} x_i p(x_i)$
  - $E[\cdot] \equiv \int_{-\infty}^{\infty} x f(x) dx$

Notation (2)

- $E[A|B]$: **conditional expectation function** (CEF) of A given B
  - $E[\cdot] \equiv \sum_{i=1}^{n} a_i p(A = a_i | B = b_i)$
  - $E[\cdot] \equiv \sum_{i=1}^{n} a_i p(A = a_i, B = b_i)/p(B = b_i)$
- $E[(A - \mu_a)]^2 \equiv V(A) \equiv \sigma_a^2$: **variance** (2nd moment)
- Note:
  - $V(A) = E[a^2] - (E[a])^2$
  - $\sqrt{V(A)} \equiv \sigma_a$ is the **standard deviation**

Association

- Recall: we cannot observe **unit treatment effect**
- We can't observe **average treatment effect** either...
- But we saw under what conditions we can estimate it
- Thus, we will focus initially on searching for ATE, which we call β (scalar)

Covariance

- Natural starting point for causal effects: **covariance**, **correlation**
- Population covariance: $Cov(Y, D) = E[(Y - E[Y])(D - E[D])]$
- Estimated covariance:

$$\widehat{Cov(Y, D)} = \frac{\sum_i (y_i - \bar{y})(d_i - \bar{d})}{n - 1} \in \mathbb{R}$$

- Limitation: no scale
- Note (for later): numerator $\sum_i (y_i - \bar{y})(d_i - \bar{d}) = \sum dy - n\bar{d}\bar{y}$
- Similarly: $\sum_i (y_i - \bar{y})^2 = \sum y^2 - n\bar{y}^2$

Correlation

- Correlation:

$$\rho \equiv Cor(Y, D) = \frac{Cov(Y, D)}{\sigma_y \sigma_d} \in [-1, 1]$$

- Limitation: no link to causality

Linear model

- Interested in the expected value of Y: $E[Y]$
- Is Y affected by D? CEF: $E[Y|D]$.
  - If $D = 0$, we write: $E[Y|D = 0]$
  - If $D = 1$, we write: $E[Y|D = 1]$
- Taking advantage of D being a 'dummy' indicator:

$$E[Y|D] = E[Y|D = 0] + D(E[Y|D = 1] - E[Y|D = 0])$$

- Define
  - $\alpha = E[Y|D = 0]$
  - $\beta = (E[Y|D = 1] - E[Y|D = 0])$
  - Then more compactly: $E[Y|D] = \alpha + \beta D$

Nota bene

$$E[Y|D] = \alpha + \beta D$$

- If $D \in \{0, 1\}$: β is $E[Y^1 - Y^0]$, or ATE
- If $D \in \mathbb{R}$: β is a slope, α an intercept (draw)
- α, β are unobserved population parameters
- Need to be estimated from the data (or not)
- Use "hat" to denote an estimate ($\hat{\beta}$) of a pop parameter
- Examples: job app, ad, RGGVY, GOTV, ...

- We are still building a model, but let's see if we can make it more friendly for the moment we encounter data...
- Recall that $E[Y|D]$ is not observable.
- Add Y (observable) on both sides gives the **linear model**

$$Y = \alpha + \beta D + (Y - E[Y|D])$$

- We often rewrite this as:

$$Y_i = \alpha + \beta D_i + \varepsilon_i$$

- No assumptions so far except for existence of $E[Y]$ (mild!)

Estimation

Estimating the model

- Many ways to estimate parameters. Even random draws! But few that meet desirable criteria
- Criterion #1: unbiasdness

$$E[\hat{\beta}] = \beta$$

- Criterion #2: consistency (prob distribution of $\hat{\beta}$ converges to β)

$$\lim_{n \to \infty} Pr[|\hat{\beta} - \beta| > \varepsilon] = 0 \forall \varepsilon > 0.$$

- Criteron #3: efficiency (minimize sampling variance $Var[\hat{\beta}]$)
- Under (sometimes restrictive) assumptions, the method of **ordinary least squares** meets these three criteria
- Let's derive OLS by hand!

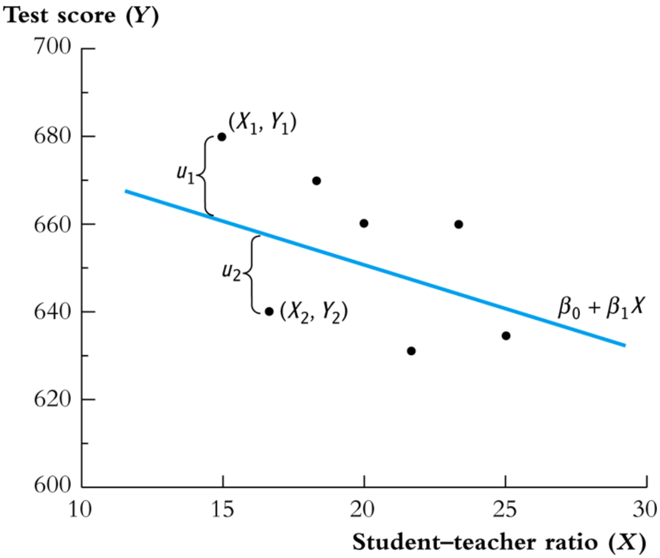
Done?

Summary

Problem:

$$\min_{\alpha,\beta} \sum_{i=1}^{n} u_i^2 = \min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - \alpha - \beta D_i)^2$$

Visually


Test score (Y)
700
680
660
640
620
600
10
15
20
25
30
Student–teacher ratio (X)
(X_1, Y_1)
(X_2, Y_2)
u_1
u_2
$\beta_0 + \beta_1 X$

Solution

Minimization problem that can be solved via FOC/SOC, which yields the "normal" equations:

$$n\hat{\alpha} + \hat{\beta}\sum x = \sum y$$
$$\hat{\alpha}\sum x + \hat{\beta}\sum x^2 = \sum xy$$

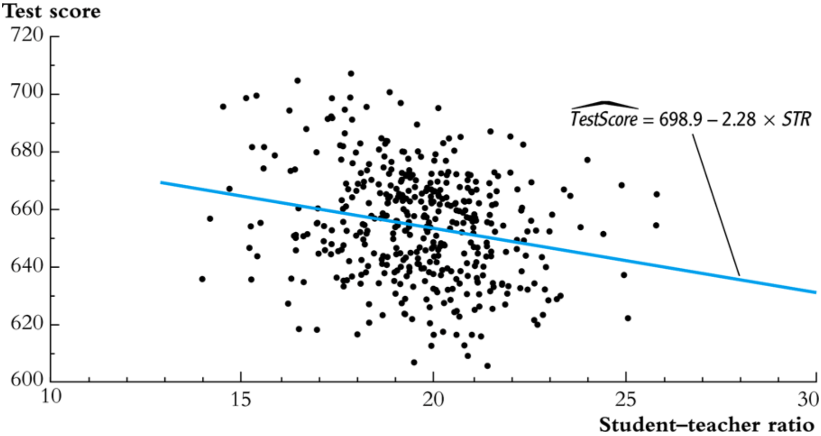
OLS estimate $\hat{\beta}$ of β (bivariate case):

$$\hat{\beta} = \frac{Cov(Y, D)}{Var(D)}$$
$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{D}$$

Interpretation

- $\hat{\beta}$ is the marginal effect of D.
- When D increases by one unit, $E[Y]$ changes by $\hat{\beta}$ units.
- Special case $D \in \{0, 1\}$: $\hat{\beta}$ is an estimate of $E[Y|D = 1] - E[Y|D = 0]$.
- $\hat{\alpha}$ is an estimate of $E[Y]$ when $D = 0$ (ie of $E[Y|D = 0]$)

Example



- $\hat{\beta} = -2.3$
- $\hat{\alpha} = 699$
- Regression equation: $\widehat{\text{Test score}} = 699 - 2.3 \cdot \text{STR}$

Example (2)

	Unique	Missing Pct.	Mean	SD	Min	Median	Max	
Donations	85	0	7075.5	5834.6	1246.0	5020.0	37015.0	
Literacy	50	0	39.3	17.4	12.0	38.0	74.0	
Commerce	84	0	42.8	25.0	1.0	42.5	86.0	
Crime_pers	85	0	19754.4	7504.7	2199.0	18748.5	37014.0	
Crime_prop	86	0	7843.1	3051.4	1368.0	7595.0	20235.0	
Clergy	85	0	43.4	25.0	1.0	43.5	86.0	

Figure 1: Table using Arel-Bundock's great *datasummary_skim*.

$$\text{Donations}_i = \alpha + \beta \text{Clergy}_i$$

For each "département" i:

Clergy: rank of nbr of priests

Donations: donations to the poor

β: effect of losing 1 rank (a bit weird – no *support*)

```
> summary(lm_robust(data = dat, formula = Donations ~ Clergy) )

Call:
lm_robust(formula = Donations ~ Clergy, data = dat)

Standard error type:  HC2

Coefficients:
            Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
(Intercept)  6210.67     1874.5  3.3133 0.001362   2483.1  9938.27 84
Clergy         19.91       34.3  0.5806 0.563092    -48.3    88.13 84

Multiple R-squared:  0.007281 , Adjusted R-squared:  -0.004537
F-statistic: 0.337 on 1 and 84 DF,  p-value: 0.5631
```

- $\hat{\alpha} = 6210$ (meaningful?)
- $\hat{\beta} = 19.9$: fewer priests leads to more donations

Properties of OLS

- **Implication 1**: $\sum u_i = 0$ and $\bar{u} = 0$.
- **Implication 2**: $Cov(D, u) = 0$.
- **Implication 3**: $\bar{\hat{y}} = \bar{y}$.
- **Implication 4**: $Cov(\hat{y}, u) = 0$.
- **Implication 5**: If D is multiplied by c, such that $z_i = cD_i$, then $\hat{\beta}_z = \hat{\beta}_d / c$ but $\hat{\alpha}_z = \hat{\alpha}_d$.
- **Implication 6**: If Y is multiplied by c, such that $z = cy_i$, then $\hat{\beta}_z = c\hat{\beta}_x$ and $\hat{\alpha}_z = c\hat{\alpha}_x$.

Questions?

References

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. London: Cambridge University Press.

Casella, George, and Roger L Berger. 2021. *Statistical Inference*. Cengage Learning.

Greene, William H. 2008. *Econometric Analysis*. 6th ed. New York: Pearson.

Stock, James H., and Mark W. Watson. 2011. *Introduction to Econometrics, 3rd Edition*. Pearson.

Verbeek, Marno. 2018. *A Guide to Modern Econometrics 5th Edition*. Wiley.