ECOM90025 Advanced Data Analysis

Tutorial 7

Zheng Fan

The University of Melbourne

Introduction

Zheng Fan

- Ph.D. student in Economics
- Email: fan.z@unimelb.edu.au
- Tutorial code and slides: github.com/zhengf1/ADA2022

Seek help?

- Ed discussion board
- Consultations: refer to Canvas for details

DiD: A short review

A famous paper by Card and Krueger (1994), investigates the following question:

- Does raising the minimum wage reduce employment?

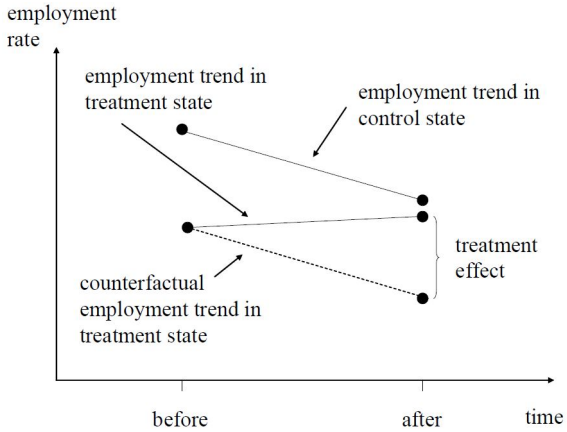
Regression: two "main" effects (time d_{Nov} and state NJ_s) and one interaction

$$Y_{ist} = \alpha + \gamma \cdot NJ_s + \lambda \cdot d_{Nov} + \delta \left(NJ_S \cdot d_{Nov}\right) + \varepsilon_{ist}$$

where the treatment is $D_{st} = NJ_s \cdot d_{Nov}$.

	$d_{Nov} = 0$	$d_{Nov} = 1$	Difference
$NJ_s = 0$	α	$\alpha + \lambda$	λ
$NJ_s = 1$	$\alpha + \gamma$	$\alpha + \gamma + \lambda + \delta$	$\lambda + \delta$
Difference	γ	$\gamma + \delta$	δ

DiD: A short review



The main identifying assumption of DiD:

- common/parallel "trends" in outcomes in treated and control groups.

DiD: Effect of a Garbage Incinerator on Housing Price

This question is Based on Kiel and McClain (1995). The data is from Wooldridge Example 13.3.

1. year: 1978 or 1981
2. age: age of house
3. agesq: age^2
4. nbh: neighborhood , 1 to 6
5. cbd: dist. to central bus. dstrct, feet
6. intst: dist. to interstate, feet
7. lintst: log(intst)
8. price: selling price
9. rooms: rooms in house
10. area: square footage of house
11. land: square footage lot
12. baths: bathrooms
13. dist: dist. from house to incinerator, feet
14. ldist: log(dist)
15. wind: perc. time wind incin. to house
16. lprice: log(price)
17. y81: =1 if year == 1981
18. larea: log(area)
19. lland: log(land)
20. y81ldist: y81*ldist
21. lintstsq: lintst^2
22. nearinc: =1 if dist <= 15840
23. y81nrinc: y81*nearinc
24. rprice: price, 1978 dollars
25. lrprice: log(rprice)

DiD: Effect of a Garbage Incinerator on Housing Price

The data description and raw data are provided to you (check Colab link). Obs: 321

1. Read the data and give proper names to each variable. Note: The raw data is not a CSV, so use 'read.table', instead of 'read.csv'.
2. Use 1981 data to estimate a linear model of 'rprice' on 'nearinc'. Why not use 'price'?

DiD: Effect of a Garbage Incinerator on Housing Price

The data description and raw data are provided to you (check Colab link). Obs: 321

1. Read the data and give proper names to each variable. Note: The raw data is not a CSV, so use 'read.table', instead of 'read.csv'.
2. Use 1981 data to estimate a linear model of 'rprice' on 'nearinc'. Why not use 'price'? rprice is price measured in 1978 dollars, which is inflation adjusted so that the 1981 and 1978 prices are comparable.
3. Use 1978 data to estimate the same model.
4. What is the treatment effect based on the previous two regression results?
5. Set up a DiD regression as in the lecture and find the treatment effect.

DiD: Effect of a Garbage Incinerator on Housing Price

	rprice		
	1981	1978	DiD
Constant	101,307.50***	82,517.23***	82,517.23***
	(3,093.03)	(2,653.79)	(2,726.91)
nearinc	−30,688.27***	−18,824.37***	−18,824.37***
	(5,827.71)	(4,744.59)	(4,875.32)
y81			18,790.29***
			(4,050.07)
nearinc:y81			−11,863.90
			(7,456.65)
Observations	142	179	321
Log Likelihood	−1,671.10	−2,095.87	−3,766.23
Akaike Inf. Crit.	3,346.19	4,195.73	7,540.46
Note:			*p<0.1; **p<0.05; ***p<0.01

DiD: Effect of a Garbage Incinerator on Housing Price

Why the difference implies a DiD?

$$\text{rprice}_{81} = \gamma_{0,81} + \gamma_{1,81}\text{nearinc}_{81} + u$$

where $\gamma_{1,81} = \overline{\text{rprice}}_{81,nr} - \overline{\text{rprice}}_{81,fr}$

$$\text{rprice}_{78} = \gamma_{0,78} + \gamma_{1,78}\text{nearinc}_{78} + u$$

where $\gamma_{1,78} = \overline{\text{rprice}}_{78,nr} - \overline{\text{rprice}}_{78,fr}$

Therefore, the DiD estimator is just

$$\hat{\delta} = \gamma_{1,81} - \gamma_{1,78} = \left(\overline{\text{rprice}}_{81,nr} - \overline{\text{rprice}}_{81,fr}\right) - \left(\overline{\text{rprice}}_{78,nr} - \overline{\text{rprice}}_{78,fr}\right)$$

Can you show this with a diagram and explain why this is a DiD?

Regression Discontinuity Design (RDD)

Let's have a competition to see who can estimate the local ATE more precisely. The true value will be revealed after the competition.

Do we have a prize for the winner???

The link in Colab can access the data.

1. y is the output
2. d is the treatment
3. r is the running variable.

A brief discussion on IV

A typical Endogeneity problem in econometrics is “omitted variable bias", where the missing variable (included in the error term) is correlated with regressors. The solutions are generally including:

- Run a RCT
- Include the extra variable(s)
- Instrumental Variables

IV requires

1. (exogeneity) $E(\epsilon_i \mid z_i) = 0$
2. (relevance) $E(x_i \mid z_i) = \pi_0 + \pi_1 z_i$ where $\pi_1 \neq 0$

The end

Thanks for your attention! 

Tutorial code and slides: github.com/zhengf1/ADA2022