

MQE: Economic Inference from Data:

Module 4: Randomized Control Trials

Claire Duquennois

6/9/2020

Causal Inference with non-random assignment:

Randomizing treatment is not always possible:

- ▶ the program or policy has already happened
- ▶ randomization is infeasible
- ▶ randomizing treatment would be unethical (ex: randomizing exposure to pollutants)

Causal Inference with non-random assignment:

With no randomized control trial we have to assume that the treatment was not randomly assigned:

- ▶ treatment will depend on observable and/or unobservable characteristics
- ▶ there are important differences between our treated and untreated units that we cannot control for
- ▶ Leaving these variables out in the error term will cause OVB.

Differences-in-differences is a way of getting around non-random assignment of a treatment.

DID: Example and intuition¹

2002: Craigslist opens a new section called “erotic services” in San Francisco:

- ▶ mostly used by sex workers to advertise and solicit clients.
- ▶ Sex workers claimed it made them safer, because they could solicit indoors from their computers and learn more about the men contacting them.
- ▶ Activists and law enforcement worried that it was facilitating sex trafficking and increasing violence against women.

Which was it? Was erotic services (ERS) making women safer, or was it placing them in harm's way?

¹This section is borrowed from [scunning.com](https://www.scunning.com)

SOS Empiricist!

This is an empirical question: What is the effect of ERS on female safety?

- ▶ The fundamental problem of causal inference strikes again!
- ▶ We can't know what effect it had because we are missing the data for the counterfactual:

$$E[\tau] = E[\underbrace{Y_{SF,2003}(D_{SF,2003} = 1)}_{\text{observed}}] - E[\underbrace{Y_{SF,2003}(D_{SF,2003} = 0)}_{\text{unobserved}}]$$

In 2003 only the first occurs, and the second is a counterfactual. So how do we proceed?

Diff-in-Diff to the rescue:

The standard differences-in-differences strategy (DiD):

- ▶ Define the intervention, D = the availability of the Craigslist ERS webpage.
- ▶ We want to know the causal effect, τ of D on Y = female murders.

Can we just compare SF murders in, say, 2003 with some other city, like Pittsburgh?

Differencing A:

....City....Outcome.....
San Francisco	$Y_{SF,2003} = \alpha_{SF} + \tau$
Pittsburgh	$Y_{P,2003} = \alpha_P$

- ▶ α_{SF} is a San Francisco fixed effect
- ▶ α_P is a Pittsburgh fixed effect.

If make a simple comparison between Pittsburgh and SF:

$$\tilde{\tau} = Y_{SF,2003} - Y_{P,2003} = \alpha_{SF} + \tau - \alpha_P.$$

Differencing A:

The simple difference is biased because of the difference in the underlying murder rates between the two cities:

$$\tilde{\tau} - \tau = \alpha_{SF} - \alpha_P.$$

Differencing B:

What if we compare SF to itself? Say in 2003 to 2001?

.....City....	..Time..Outcome.....
San Francisco	Before	$Y_{SF,2001} = \alpha_{SF}$
	After	$Y_{SF,2003} = \alpha_{SF} + \lambda_{03} + \tau$

Again, this doesn't lead to an unbiased estimate of τ since:

$$\tilde{\tau} = \alpha_{SF} + \lambda_{03} + \tau - \alpha_{SF} = \lambda_{03} + \tau$$

We eliminated the city fixed effect but not the changes in the murder rate over time which will bias my estimate:

$$\tilde{\tau} - \tau = \lambda_{03}$$

How can I identify and control for these time effects?

Differencing $A+B=$ Diff-in-Diff

Combining the two approaches to eliminate both the city effects and the time effects:

... City...	..Time..Outcome.....	1st Diff	2nd Diff
SF	Before	$Y_{SF,2001} = \alpha_{SF}$		
	After	$Y_{SF,2003} = \alpha_{SF} + \lambda_{03} + \tau$	$\lambda_{03} + \tau$	τ
Pittsburgh	Before	$Y_{P,2001} = \alpha_P$		
	After	$Y_{P,2003} = \alpha_P + \lambda_{03}$	λ_{03}	

The idea

Sometimes treatment and control group outcomes move in parallel in the absence of treatment.

When they do, the divergence of a post-treatment path from the trend established by a comparison group may signal a treatment effect.

The mechanics

Difference-in-differences can be implemented as follows:

- 1) Compute the difference in the mean outcome variable Y in the post treatment period ($t = 1$) and the before treatment period ($t = 0$) for the control group C :

$$\bar{Y}_{C,1} - \bar{Y}_{C,0} = \Delta \bar{Y}_C$$

\Rightarrow allows us to cancel out the control group fixed effect and identify the time fixed effect since

$$\bar{Y}_{C,1} - \bar{Y}_{C,0} = \alpha_C + \lambda_1 - \alpha_c = \lambda_1 = \Delta \bar{Y}_C$$