Longitudinal Data Analysis

methods@manchester summer school

Day 4 | morning session

Thiago R. Oliveira

Lecturer in Quantitative Criminology, University of Manchester

30/06—04/07

Today

Reverse causality and reciprocal relationships

- ⇝ Causal inference and the potential outcomes framework
- ⇝ The fundamental problem of causal inference
- ⇝ Leveraging longitudinal data
- ⇝ The difference-in-differences analysis
- ⇝ Estimating CLPMs using `lavaan`

Potential outcomes framework

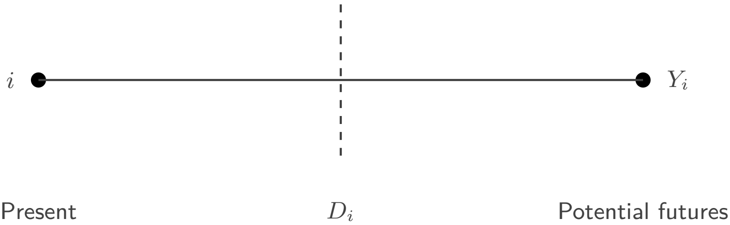
Causal inference and the potential outcomes framework

Goal in causal inference is to assess the causal effect of a treatment/exposure on some outcome

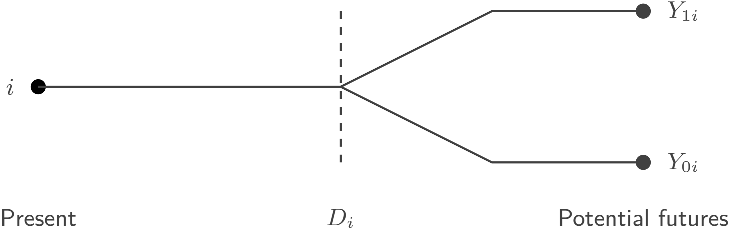
- Does raising the minimum wage reduce employment?
- Does housing assistance reduce homelessness?
- Does smoking cause lung cancer?
- Does voting by mail increase voter turnout?
- Does exposure to misinformation reduce political trust??
- ...

Causal inference and the potential outcomes framework

Causal inference and the potential outcomes framework



Causal inference and the potential outcomes framework



Causal inference and the potential outcomes framework

Y_i: Observed outcome variable of interest for unit i

Potential outcomes

Y_{0i} and Y_{1i}: Potential outcomes for unit i

$$Y_{\cdot i} = \begin{cases} Y_{1i} & \text{Potential outcome for unit } i \text{ with treatment} \\ Y_{0i} & \text{Potential outcome for unit } i \text{ without treatment} \end{cases}$$

D_i: Indicator of treatment intake for *unit* i

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment} \\ 0 & \text{otherwise.} \end{cases}$$

Definition of causal effect

$$\delta_i = Y_{1i} - Y_{0i}$$

Fundamental problem of causal inference

$\rightsquigarrow$ We cannot observe both potential outcomes for the same unit i!

Causal inference and the potential outcomes framework

Y_i: Observed outcome variable of interest for unit i

Potential outcomes

Y_{0i} and Y_{1i}: Potential outcomes for unit i

$$Y_{\cdot i} = \begin{cases} Y_{1i} & \text{Potential outcome for unit } i \text{ with treatment} \\ Y_{0i} & \text{Potential outcome for unit } i \text{ without treatment} \end{cases}$$

D_i: Indicator of treatment intake for *unit* i

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment} \\ 0 & \text{otherwise.} \end{cases}$$

Definition of causal effect

$$\delta_i = Y_{1i} - Y_{0i}$$

Fundamental problem of causal inference

$\rightsquigarrow$ We cannot observe both potential outcomes for the same unit i!

Causal inference and the potential outcomes framework

Y_i: Observed outcome variable of interest for unit i

Potential outcomes

Y_{0i} and Y_{1i}: Potential outcomes for unit i

$$Y_{\cdot i} = \begin{cases} Y_{1i} & \text{Potential outcome for unit } i \text{ with treatment} \\ Y_{0i} & \text{Potential outcome for unit } i \text{ without treatment} \end{cases}$$

D_i: Indicator of treatment intake for *unit* i

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment} \\ 0 & \text{otherwise.} \end{cases}$$

Definition of causal effect

$$\delta_i = Y_{1i} - Y_{0i}$$

Fundamental problem of causal inference

$\rightsquigarrow$ We cannot observe both potential outcomes for the same unit i!

Causal inference and the potential outcomes framework

Y_i: Observed outcome variable of interest for unit i

Potential outcomes

Y_{0i} and Y_{1i}: Potential outcomes for unit i

$$Y_{\cdot i} = \begin{cases} Y_{1i} & \text{Potential outcome for unit } i \text{ with treatment} \\ Y_{0i} & \text{Potential outcome for unit } i \text{ without treatment} \end{cases}$$

D_i: Indicator of treatment intake for *unit* i

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment} \\ 0 & \text{otherwise.} \end{cases}$$

Definition of causal effect

$$\delta_i = Y_{1i} - Y_{0i}$$

Fundamental problem of causal inference

$\rightsquigarrow$ We cannot observe both potential outcomes for the same unit i!

Causal inference and the potential outcomes framework

Randomisation solves the problem!

Logic of randomised control trials

- ⇝ Randomly divide a sample in two groups
- ⇝ Because this was random, both groups are *on average* the same
- ⇝ Then apply the treatment/exposure to one group (the treatment group), but not the other (control group)
- ⇝ Because the exposure happened after the treatment assignment, the only difference between the two groups is the treatment/exposure
- ⇝ Therefore, any subsequently observed differences are attributable to the treatment/exposure
- ⇝ We randomisation, we can thus find the **average treatment effect**

Causal inference and the potential outcomes framework

Randomisation solves the problem!

Logic of randomised control trials

- Randomly divide a sample in two groups
- Because this was random, both groups are *on average* the same
- Then apply the treatment/exposure to one group (the treatment group), but not the other (control group)
- Because the exposure happened after the treatment assignment, the only difference between the two groups is the treatment/exposure
- Therefore, any subsequently observed differences are attributable to the treatment/exposure
- We randomisation, we can thus find the **average treatment effect**

Causal inference and the potential outcomes framework

Randomisation solves the problem!

Logic of randomised control trials

- ⇝ Randomly divide a sample in two groups
- ⇝ Because this was random, both groups are *on average* the same
- ⇝ Then apply the treatment/exposure to one group (the treatment group), but not the other (control group)
- ⇝ Because the exposure happened after the treatment assignment, the only difference between the two groups is the treatment/exposure
- ⇝ Therefore, any subsequently observed differences are attributable to the treatment/exposure
- ⇝ We randomisation, we can thus find the **average treatment effect**

Causal inference and the potential outcomes framework

What if we cannot conduct an experiment?

- ⇝ Randomised Experiments
- ⇝ Observational Studies
    - Selection on observables
        - Regression
        - Matching
        - Weighting
    - Selection on unobservables
        - Difference-in-Differences and synthetic control
        - Instrumental Variables
        - Regression Discontinuity Designs

Causal inference and the potential outcomes framework

- $\rightsquigarrow$ Causality is defined by potential outcomes, not by realised (observed) outcomes
- $\rightsquigarrow$ Observed association is neither necessary nor sufficient for causality
- $\rightsquigarrow$ Estimation of causal effects of a treatment (usually) starts with studying the assignment mechanism
- $\rightsquigarrow$ The goal is to mimic the features of a randomised experiment even if we don't have one
- $\rightsquigarrow$ When we don't have an RCT, our ability to make causal inferences often relies on making untestable assumptions about the assignment mechanism

$\Rightarrow$ Now let's see how we can leverage panel data to make causal inferences!

Difference-in-differences

Intuition of the difference-in-differences estimator

$\Rightarrow$ What if we use **time** in our favour?

- $\leadsto$ Collect data on Y at two points in time: **before** and **after** the treatment/exposure/policy intervention
- $\leadsto$ Analyse the extent to which Y **changes** in units that received the treatment
- $\leadsto$ Analyse the extent to which Y **changes** in units that did NOT receive the treatment
- $\leadsto$ Compare the two **changes**

Intuition of the difference-in-differences estimator

$\Rightarrow$ What if we use **time** in our favour?

- $\rightsquigarrow$ Collect data on Y at two points in time: **before** and **after** the treatment/exposure/policy intervention
- $\rightsquigarrow$ Analyse the extent to which Y **changes** in units that received the treatment
- $\rightsquigarrow$ Analyse the extent to which Y **changes** in units that did NOT receive the treatment
- $\rightsquigarrow$ Compare the two **changes**

Intuition of the difference-in-differences estimator

$\Rightarrow$ What if we use **time** in our favour?

- $\rightsquigarrow$ Collect data on Y at two points in time: **before** and **after** the treatment/exposure/policy intervention
- $\rightsquigarrow$ Analyse the extent to which Y **changes** in units that received the treatment
- $\rightsquigarrow$ Analyse the extent to which Y **changes** in units that did NOT receive the treatment
- $\rightsquigarrow$ Compare the two **changes**

Intuition of the difference-in-differences estimator

Some conceptual clarification to make our lives easier

- Variation **between** units: **difference**
- Variation **within** units (over time): **changes**

⇒ We want to estimate the **difference in changes**
or (*difference-in-differences*)

- The difference between (a) changes in Y before and after the intervention among treated units and (b) changes in Y before and after the intervention among non-treated units is the **causal effect**!

(under some assumptions regarding those changes. . . Let's dive into it)

Intuition of the difference-in-differences estimator

Some conceptual clarification to make our lives easier

- ⇝ Variation **between** units: **difference**
- ⇝ Variation **within** units (over time): **changes**

⇒ We want to estimate the **difference in changes**
or (*difference-in-differences*)

- ⇝ The difference between (a) changes in Y before and after the intervention among treated units and (b) changes in Y before and after the intervention among non-treated units is the **causal effect**!

(under some assumptions regarding those changes. . . Let's dive into it)

Intuition of the difference-in-differences estimator

Some conceptual clarification to make our lives easier

- ⇝ Variation **between** units: **difference**
- ⇝ Variation **within** units (over time): **changes**

⇒ We want to estimate the **difference in changes**
or *(difference-in-differences)*

- ⇝ The difference between (a) changes in Y before and after the intervention among treated units and (b) changes in Y before and after the intervention among non-treated units is the **causal effect**!

(under some assumptions regarding those changes. . . Let's dive into it)

Intuition of the difference-in-differences estimator

Some conceptual clarification to make our lives easier

- ⇝ Variation **between** units: **difference**
- ⇝ Variation **within** units (over time): **changes**

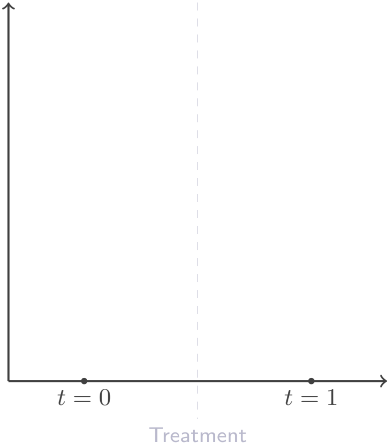
⇒ We want to estimate the **difference in changes**
or *(difference-in-differences)*

- ⇝ The difference between (a) changes in Y before and after the intervention among treated units and (b) changes in Y before and after the intervention among non-treated units is the **causal effect**!

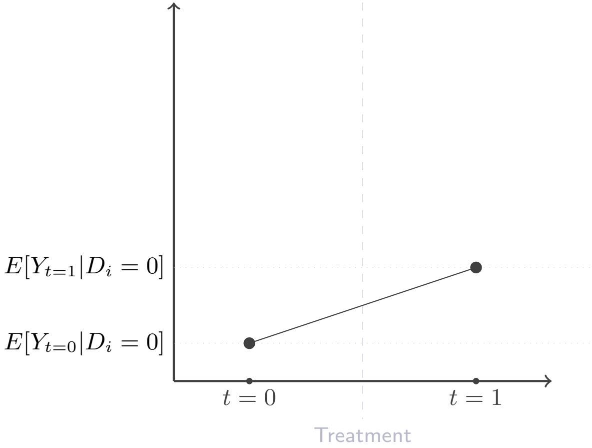
(under some assumptions regarding those changes. . . Let's dive into it)

The two-period setup

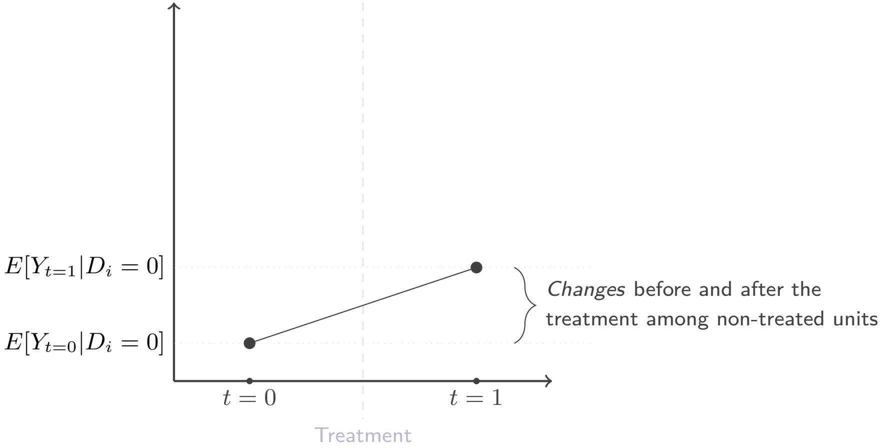
Difference-in-differences setup



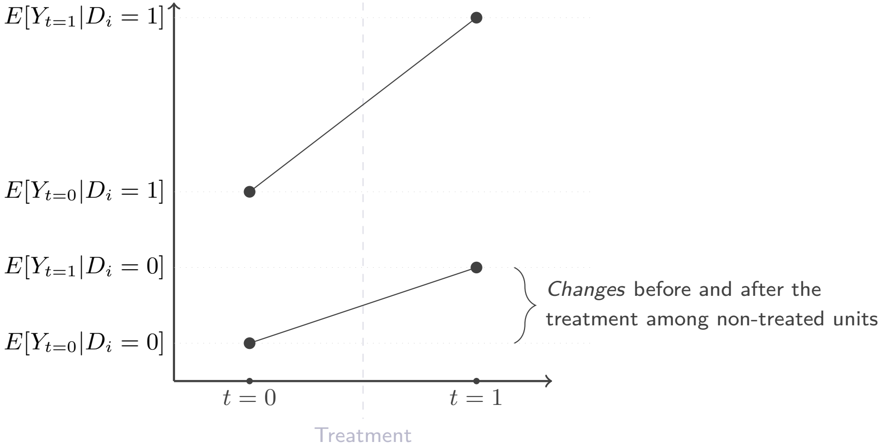
Difference-in-differences setup



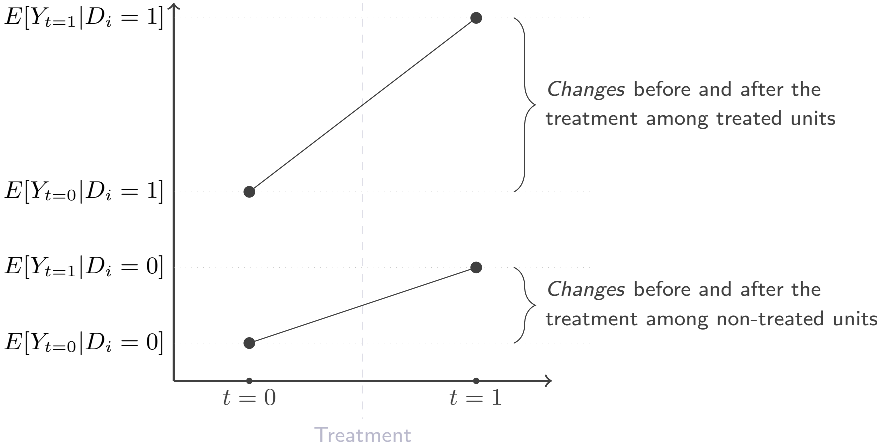
Difference-in-differences setup



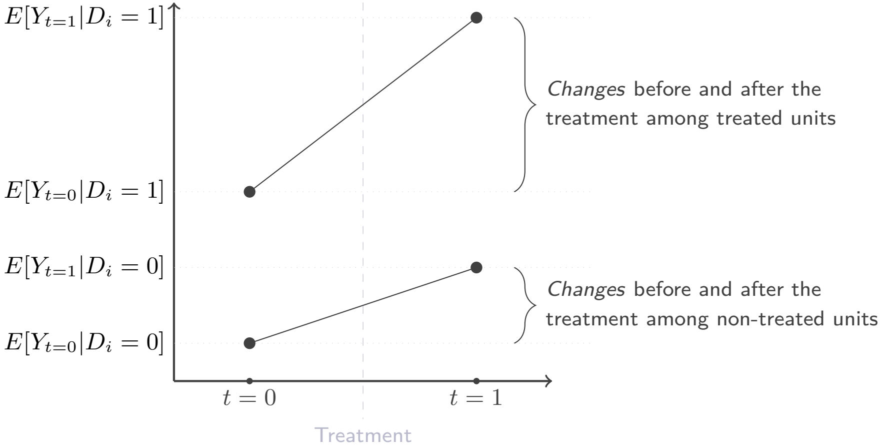
Difference-in-differences setup



Difference-in-differences setup



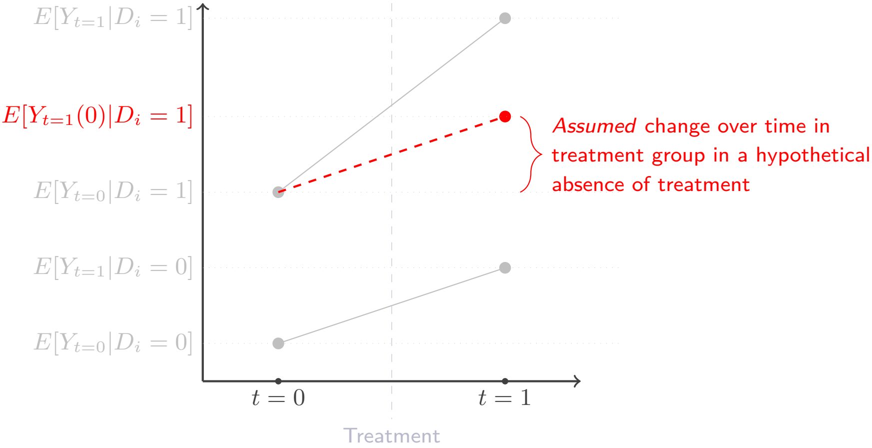
Difference-in-differences setup



⇝ **Problem**: Missing potential outcomes: $E[Y_{i,t=1}(0)|D_i = 1]$ and $E[Y_{i,t=1}(1)|D_i = 0]$

Difference-in-differences setup

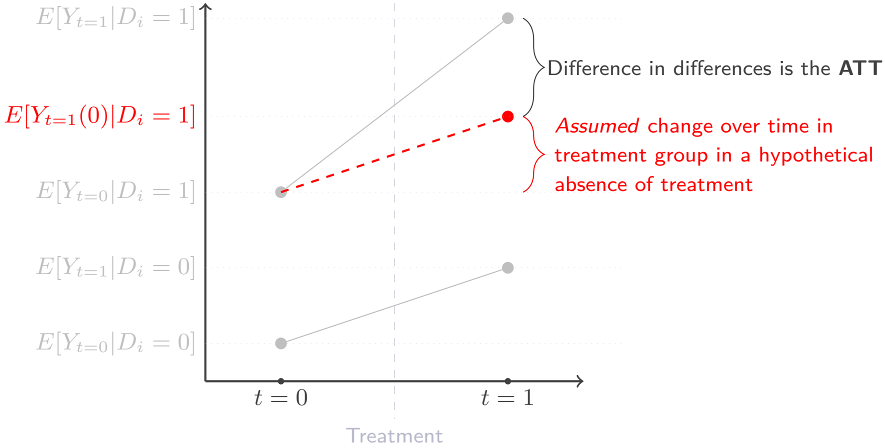
Strategy: Use the change in the control group to *assume* $E[Y_{t=1}(0)|D_i = 1]$



Assumption: Trend over time would be the same for treatment and control

Difference-in-differences setup

Strategy: Use the change in the control group to *assume* $E[Y_{t=1}(0)|D_i = 1]$



Assumption: Trend over time would be the same for treatment and control

Identification assumption

Parallel trends

- $\rightsquigarrow$ Had the treated units not received the treatment, they would have followed the same trend as the control units

Difference-in-differences estimator

Difference in changes:

$$\delta_{ATT} = \Big\{\text{Changes in treatment group before and after treatment}\Big\} - \Big\{\text{Changes in control group before and after treatment}\Big\}$$

Threats to validity

Non-parallel trends

$\rightsquigarrow$ Very critical assumption: treatment units have similar trends to control units in the absence of treatment

$\rightsquigarrow$ Fundamental problem of causal inference: we cannot observe potential outcome under the control condition for treated units in the post-treatment period

$\Rightarrow$ What can we do?

(more on that later...)

- Careful assessment: is assuming parallel trends plausible?
- Estimate treatment effects at different time points (placebo tests)

Thank you!

thiago.oliveira@manchester.ac.uk

ThiagoROliveira.com

@oliveiratr.bsky.social