Machine Learning for econometrics

Event studies: Causal methods for pannel data

Authors

February, 11th, 2025

Motivation

Setup: event studies

Estimation of the effect of a treatment when data is

Aggregated: eg. country-level data such as employment rate, GDP, etc.

Setup: event studies

Estimation of the effect of a treatment when data is

Aggregated: eg. country-level data such as employment rate, GDP, etc.

Longitudinal: eg. multiple time periods or repeated cross-sections

Setup: event studies

Estimation of the effect of a treatment when data is

Aggregated: eg. country-level data such as employment rate, GDP, etc.

Longitudinal: eg. multiple time periods or repeated cross-sections

With multiple units: eg. multiple countries, firms, regions.

Setup: event studies

Estimation of the effect of a treatment when data is

Aggregated: eg. country-level data such as employment rate, GDP, etc.

Longitudinal: eg. multiple time periods or repeated cross-sections

With multiple units: eg. multiple countries, firms, regions.

Staggered adoption of the treatment: eg. different countries adopt a policy at different times.

Setup: event studies

Estimation of the effect of a treatment when data is

Aggregated: eg. country-level data such as employment rate, GDP, etc.

Longitudinal: eg. multiple time periods or repeated cross-sections

With multiple units: eg. multiple countries, firms, regions.

Staggered adoption of the treatment: eg. different countries adopt a policy at different times.

This setup is known as: panel data, event studies, longitudinal data, time-series data.

Examples of event studyies for policy question

Setup: event studies are quasi-experiment

- Quasi-experiment: a situation where the treatment is not randomly assigned by the researcher but by nature or society.
- Should introduces some randomness in the treatment assignment: enforcing treatment exogeneity, ie. ignorability (ie. unconfoundedness).

Today: Three quasi-experimental designs for event studies

- The simple method of difference-in-differences with a strong assumption called paralled trend
- Synthetic control method: a balancing method (think to propensity score matching)
- Conditional DID: a doubly robust method combining outcomes and propensity score models

Table of contents

Reminder on difference-in-differences

History

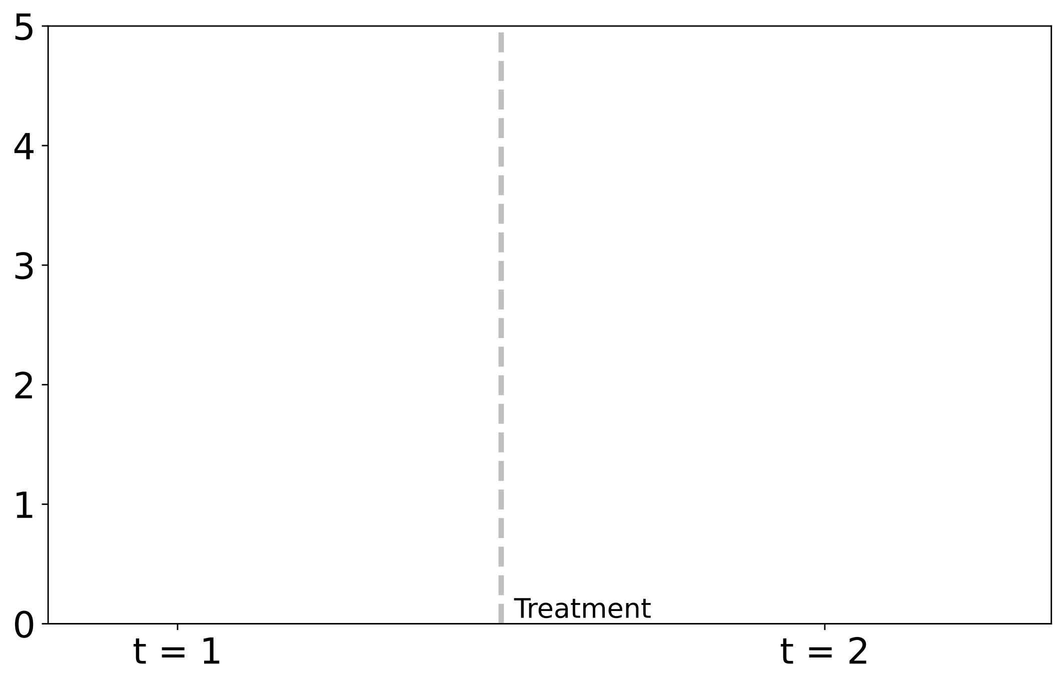
- First documented example (though not formalized): John Snow showing how cholera spread through the water in London (Snow, 1855)[1]
- Modern usage introduced formally by (Ashenfelter, 1978), applied to labor economics

Idea

- Contrast the temporal effect of the treated unit with the control unit temporal effect:
- The difference between the two differences is the treatment effect

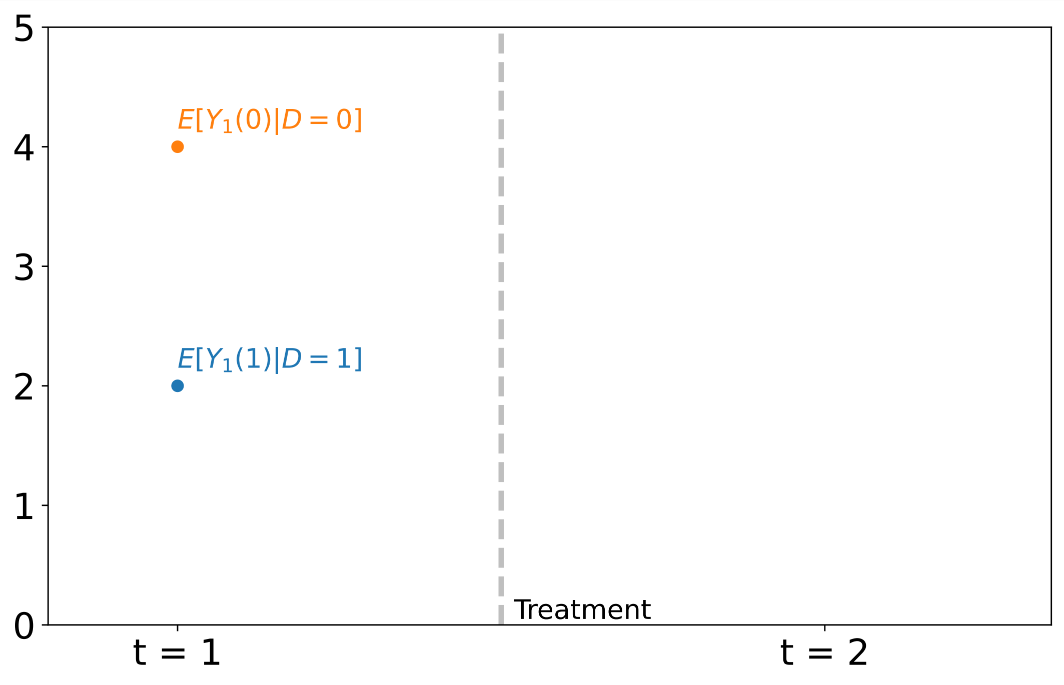
[1] Good description: https://mixtape.scunning.com/09-difference_in_differences#john-snows-cholera-hypothesis

Two period of times: t=1, t=2



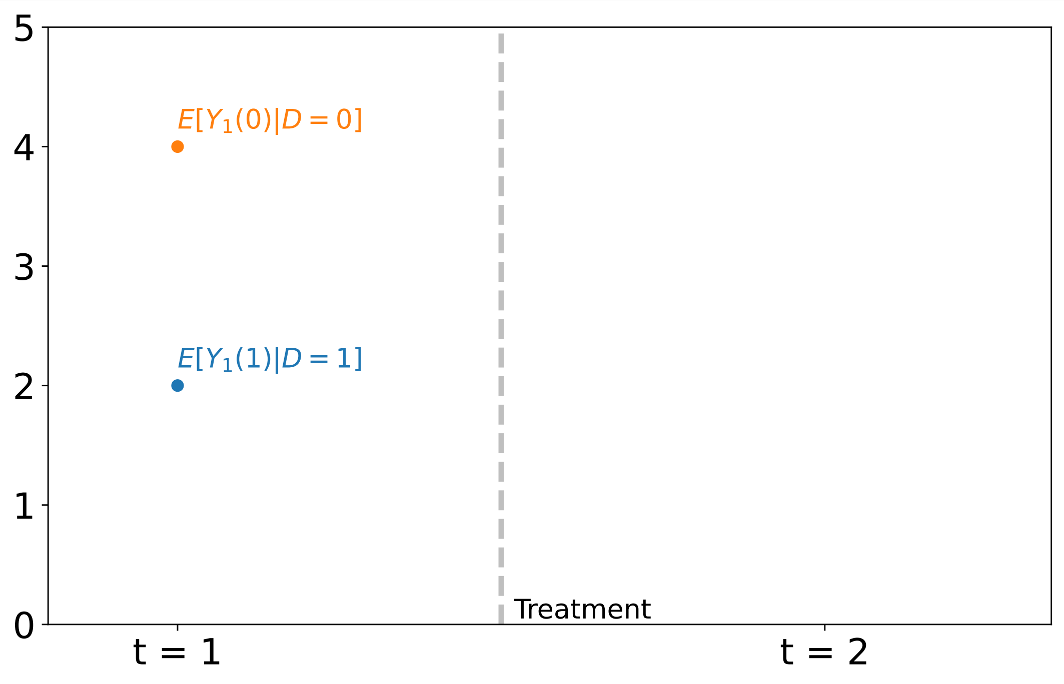
Difference-in-differences framework

Potential outcomes: $Y_t(d)$ where $d = \{0, 1\}$ is the treatment at period 2



Difference-in-differences framework

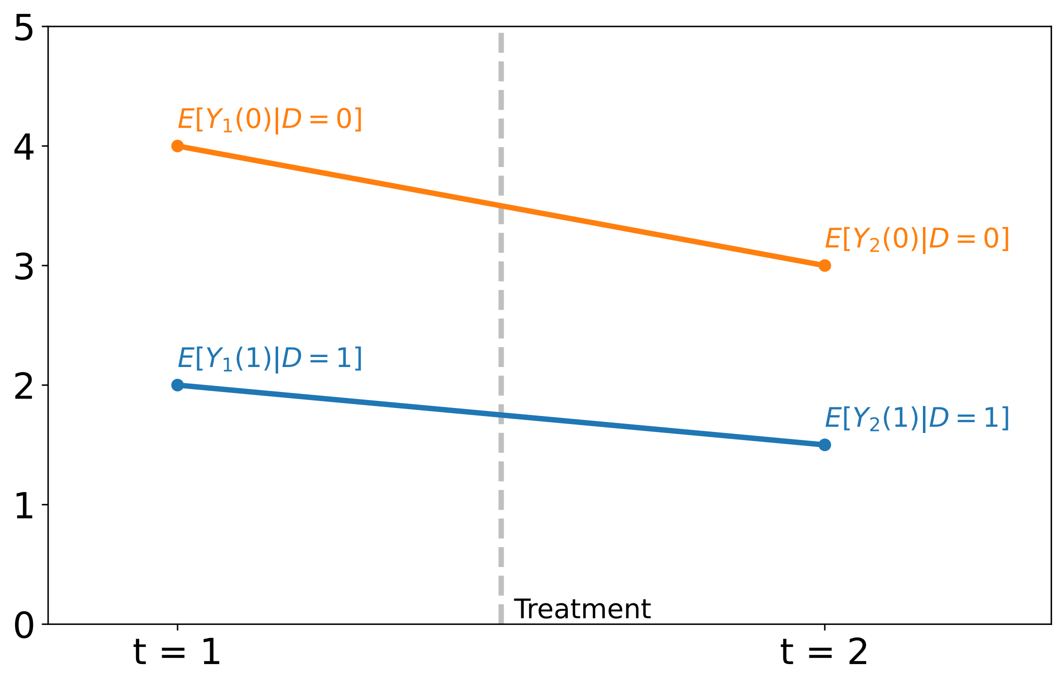
Potential outcomes: $Y_t(d)$ where $d = \{0, 1\}$ is the treatment at period 2



⚠️ $\mathbb{E}[Y_1(1)] = \mathbb{E}[Y_1(1) \; |D=1]\mathbb{P}(D=1) + \mathbb{E}[Y_1(1) \; |D=0]\mathbb{P}(D=0)$
but we only observe $\mathbb{E}[Y_1(1) \; |D=1]$

Difference-in-differences framework

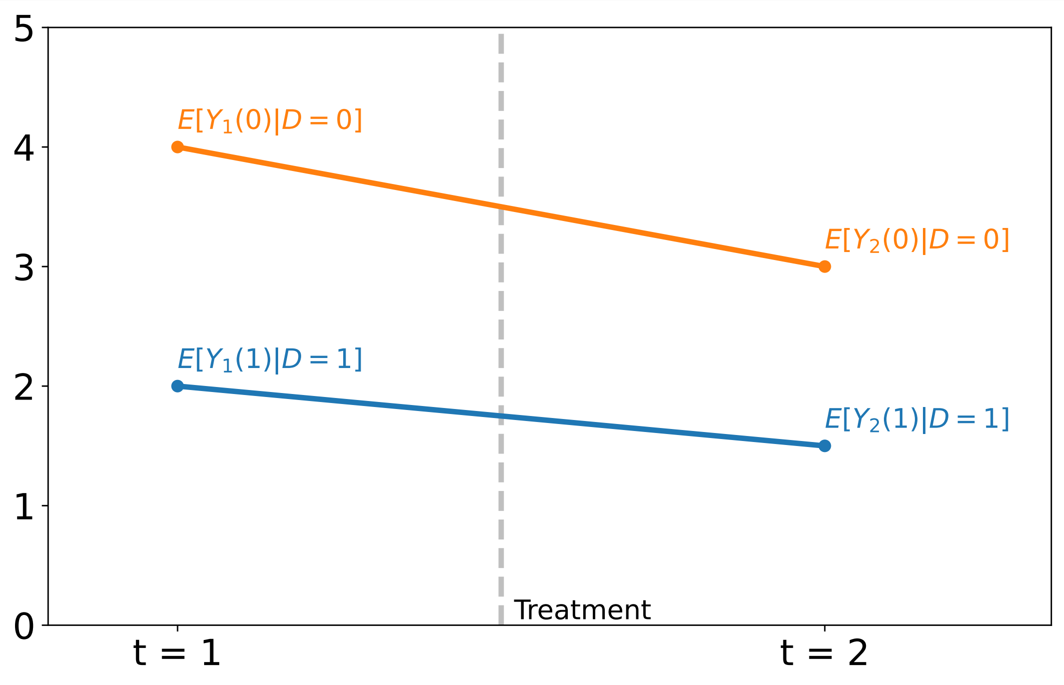
Our target is the average treatment effect on the treated (ATT)



$$\tau_{\text{ATT}} = \mathbb{E}[Y_2(1)|\ D = 1] - \mathbb{E}[Y_2(0)|\ D = 1]$$

Difference-in-differences framework

Our target is the average treatment effect on the treated (ATT)

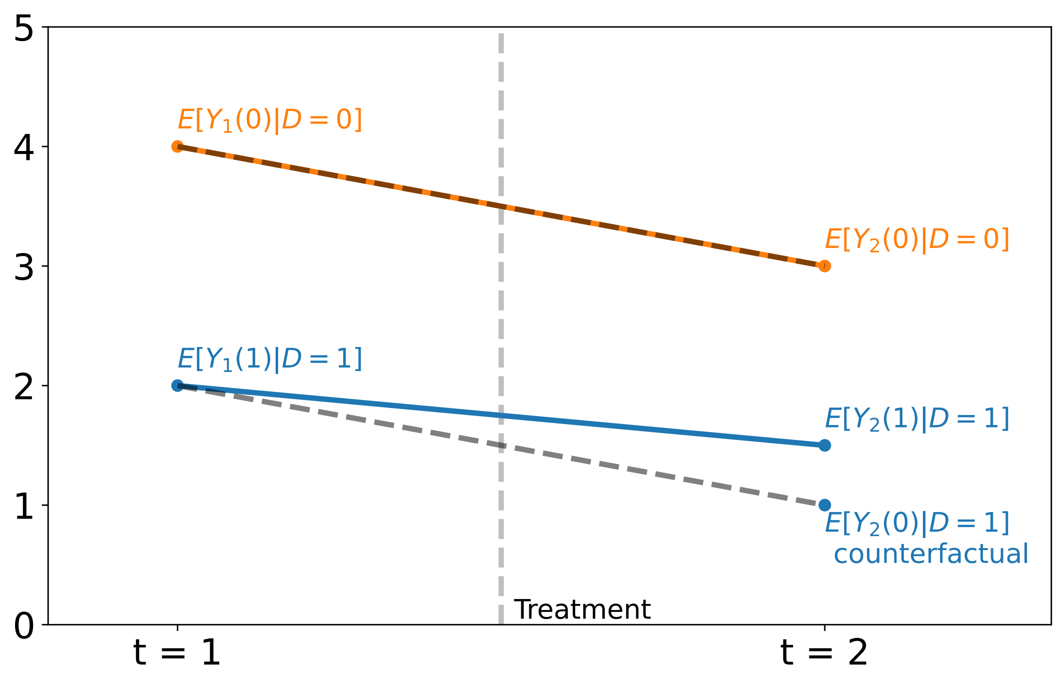


$$\tau_{\mathrm{ATT}} = \underbrace{[Y_2(1)|\ D=1]}_{\text{treated outcome for t=2}} - \underbrace{\mathbb{E}[Y_2(0)|\ D=1]}_{\text{unobserved counterfactual}}$$

Difference-in-differences framework

First assumption, parallel trends

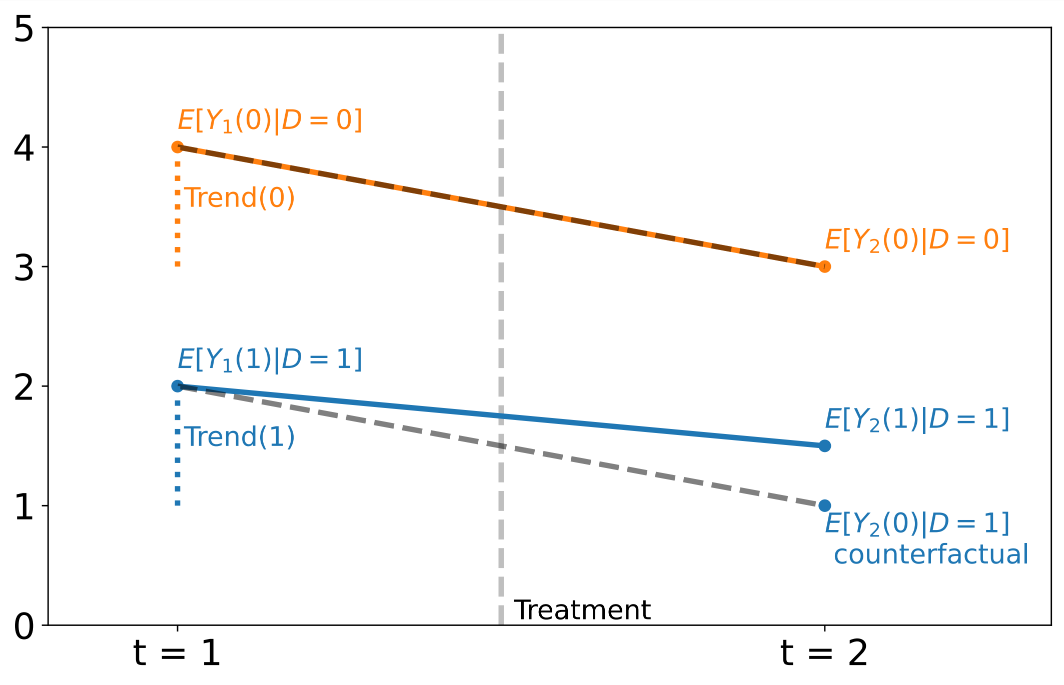
$$\mathbb{E}[Y_2(0) - Y_1(0) \mid D = 1] = \mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]$$



Difference-in-differences framework

First assumption, parallel trends

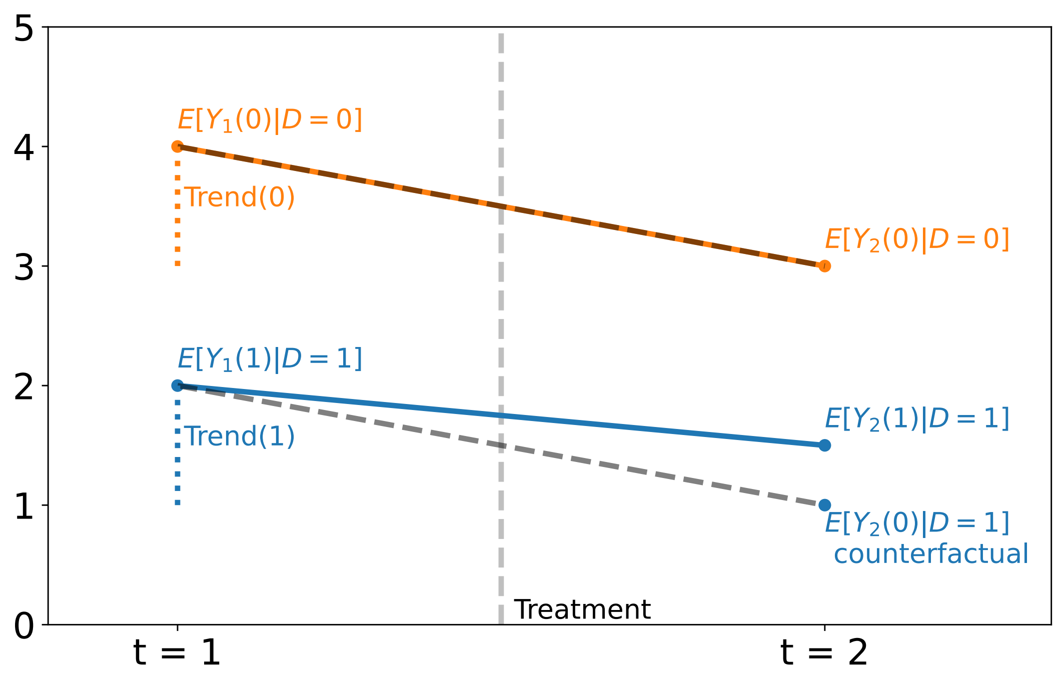
$$\underbrace{[Y_2(0) - Y_1(0) \mid D = 1]}_{\color{blue}{\text{Trend}(1)}} = \underbrace{\mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]}_{\color{orange}{\text{Trend}(0)}}$$



Difference-in-differences framework

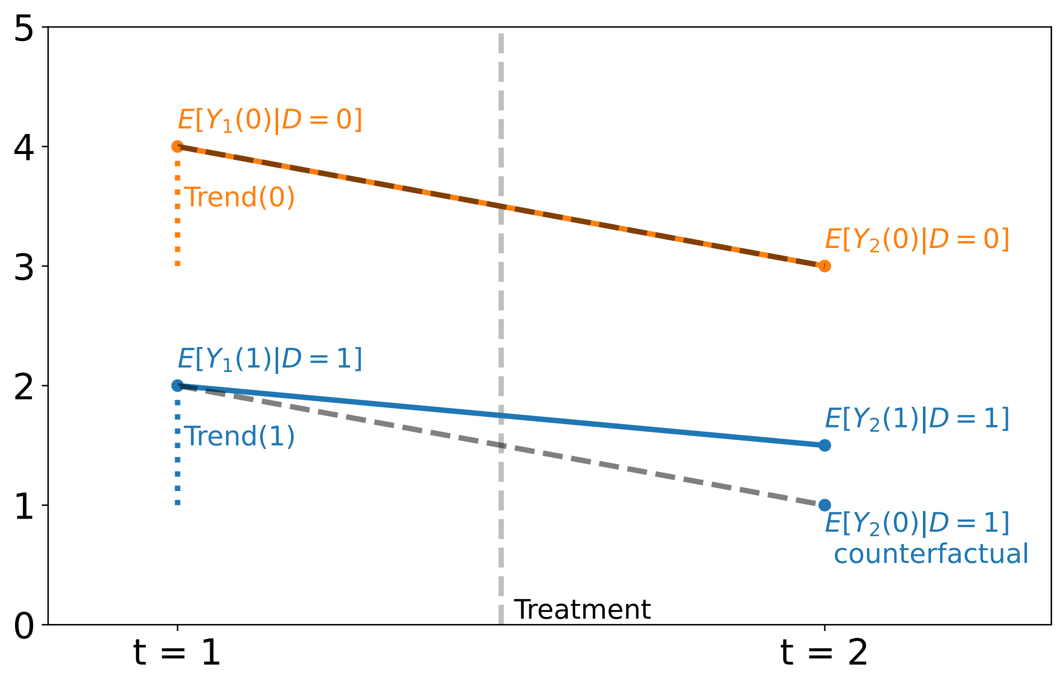
First assumption, parallel trends

$$\mathbb{E}[Y_2(0) \mid D = 1] = \mathbb{E}[Y_1(0) \mid D = 1] + \mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]$$



First assumption, parallel trends

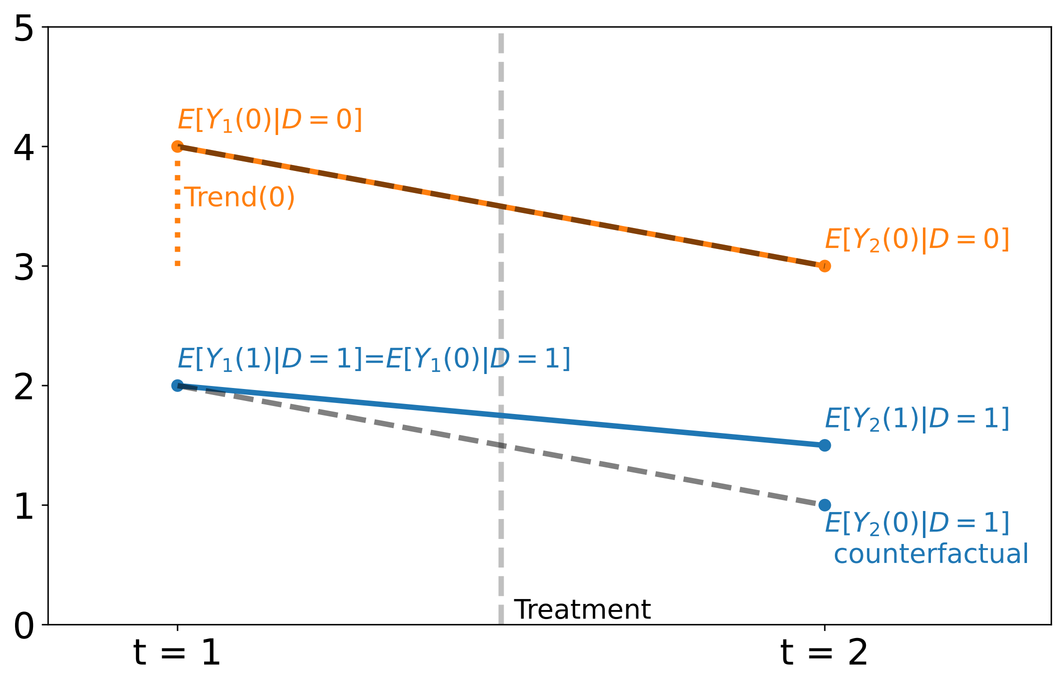
$$\mathbb{E}[Y_2(0) \mid D = 1] = \underbrace{[Y_1(0) \mid D = 1]}_{\text{unobserved counterfactual}} + \mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]$$



Difference-in-differences framework

Second assumption, no anticipation of the treatment

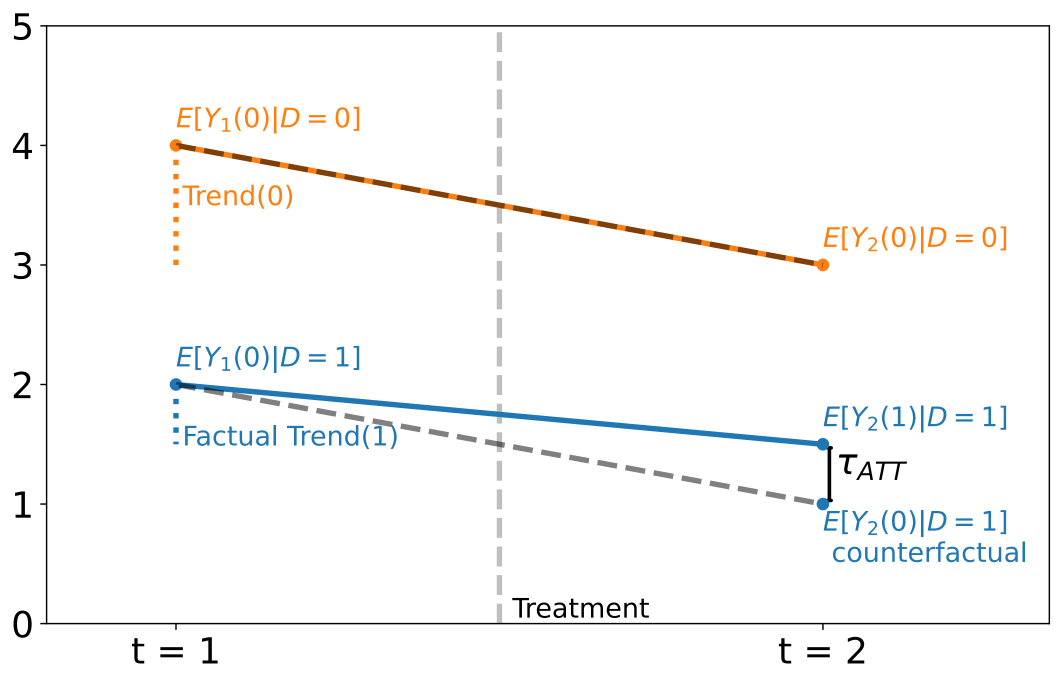
$$E[Y_1(1)|D=1] = E[Y_1(0)|D=1]$$



Difference-in-differences framework: identification of ATT

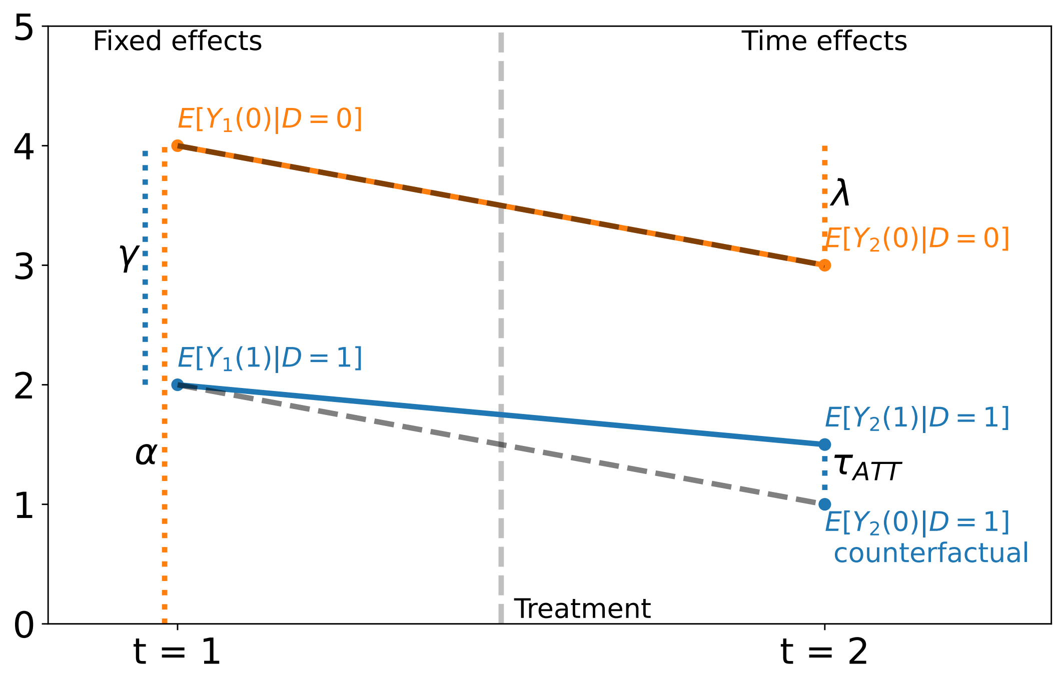
$$\tau_{\text{ATT}} = \mathbb{E}[Y_2(1)|\ D = 1] - \mathbb{E}[Y_2(0)|\ D = 1]$$

$$= \underbrace{\mathbb{E}[Y_2(1)|\ D = 1] - \mathbb{E}[Y_1(0)|D = 1]}_{\color{blue}{\text{Factual Trend}(1)}} - \underbrace{\mathbb{E}[Y_2(0)|D = 0] - \mathbb{E}[Y_1(0)|D = 0]}_{\color{orange}{\text{Trend}(0)}}$$



Estimation: link with two way fixed effect (TWFE)

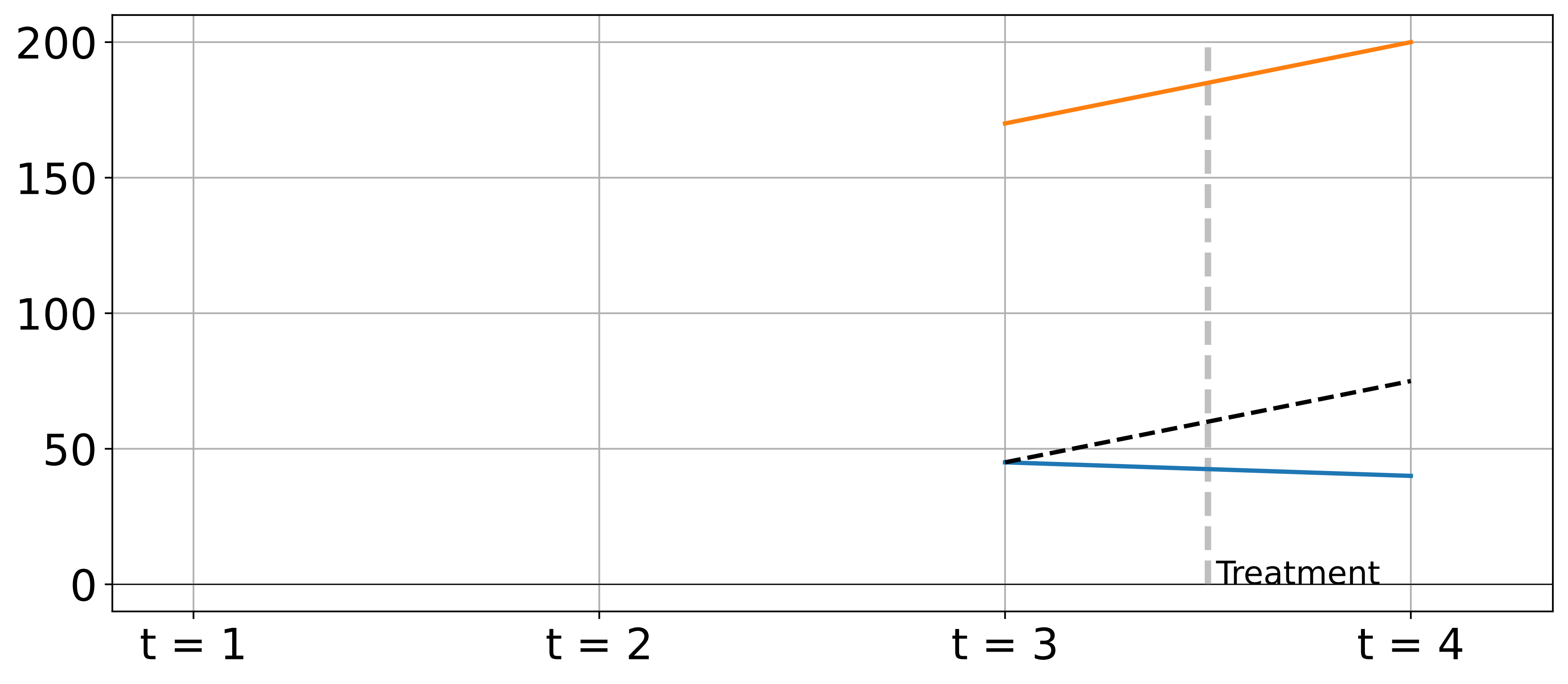
$$Y = \alpha + \gamma D + \lambda \mathbb{1}(t = 2) + \tau_{\text{ATT}} D \mathbb{1}(t = 2)$$



⚠️ Mechanic link working only with assumptions

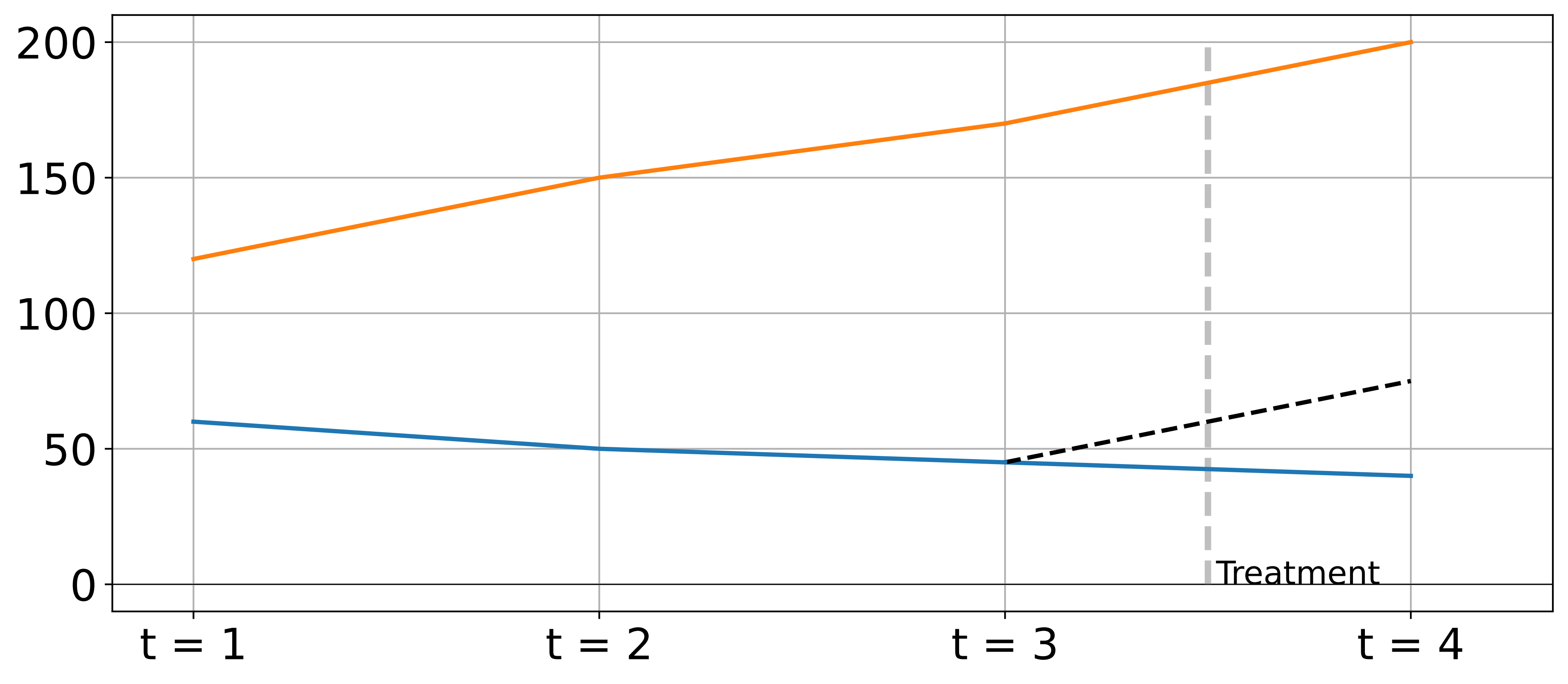
Failure of the parallel trend assumption

Seems like the treatment decreases the outcome!



Failure of the parallel trend assumption

Oups…



DID estimator for more than two time units

Target estimand: sample average treatment effect on the treated (SATT)

$$\tau_{\text{SATT}} = \frac{1}{|\{i : D_i = 1\}|} \sum_{i:D_i=1} \frac{1}{T-H} \sum_{t=H+1}^{T} Y_{it}(1) - Y_{it}(0)$$

DID estimator

$$\widehat{\tau_{\text{DID}}} = \frac{1}{|\{i : D_i = 1\}|} \sum_{i:D_i=1} \left[\frac{1}{T-H} \sum_{t=H+1}^{T} Y_{it} - \frac{1}{H} \sum_{t=1}^{H} Y_{it}\right] - \frac{1}{|\{i : D_i = 0\}|} \sum_{i:D_i=0} \left[\frac{1}{T-H} \sum_{t=H+1}^{T} Y_{it} - \frac{1}{H} \sum_{t=1}^{H} Y_{it}\right]$$

Assumption

No anticipation of the treatment: $Y_{it}(0) = Y_{it}(1) \forall t = 1, ..., H$.

Parallel trend: $\mathbb{E}[Y_{it}(0, \infty) - Y_{i1}(0, \infty)] = \beta_t, t = 2, ..., T$.

See (Wager, 2024) for a clear proof of consistancy.

DID: Take-away

- Extremely common in economics
- Very strong assumptions: parallel trends and no anticipation
- Can be extended to (Wager, 2024):
  - more than two time periods: exact same formulation
  - staggered adoption of the treatment: a bit more complex
- Does not account for heterogeneity of treatment effect over time

Synthetic Controls

Synthetic Controls

Introduced in (Abadie & Gardeazabal, 2003) and (Abadie et al., 2010), well described in (Abadie, 2021)

- Estimates the effect of a treatment on a single unit
- The treatment unit is compared to a weighted average of control units
- The weights are chosen to minimize the difference between the treated unit and the synthetic control

Example

- What is the effect of taxes on sugar-based product consumption (Puig-Codina et al., 2021)
- Review for epidemiology (Bonander et al., 2021).

Conditional difference-in-differences

Time-series modelisation: methods without a control group

Interrupted Time Series

State space models

Take-away

Python hands-on

To your notebooks 🧑‍💻 !

- url: https://github.com/strayMat/causal-ml-course/tree/main/notebooks

Bibliography

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. Journal of Economic Literature, 59(2), 391–425.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. American Economic Review, 93(1), 113–132.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Journal of the American Statistical Association, 105(490), 493–505.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. The Review of Economics and Statistics, 47–57.

Bonander, C., Humphreys, D., & Degli Esposti, M. (2021). Synthetic control methods for the evaluation of single-unit interventions in epidemiology: a tutorial. American Journal of Epidemiology, 190(12), 2700–2711.

Bibliography

Puig-Codina, L., Pinilla, J., & Puig-Junoy, J. (2021). The impact of taxing sugar-sweetened beverages on cola purchasing in Catalonia: an approach to causal inference with time series cross-sectional data. The European Journal of Health Economics, 22(1), 155–168.

Snow, J. (1855). On the mode of communication of cholera. John Churchill.

Wager, S. (2024,). Causal inference: A statistical learning approach. preparation.