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

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.

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.
- A convincing quasi-experiment introduces a certain amount of randomness in the treatment assignment (sometimes called exogeneity): it enforces the ignorability assumption (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

Difference-in-differences

History

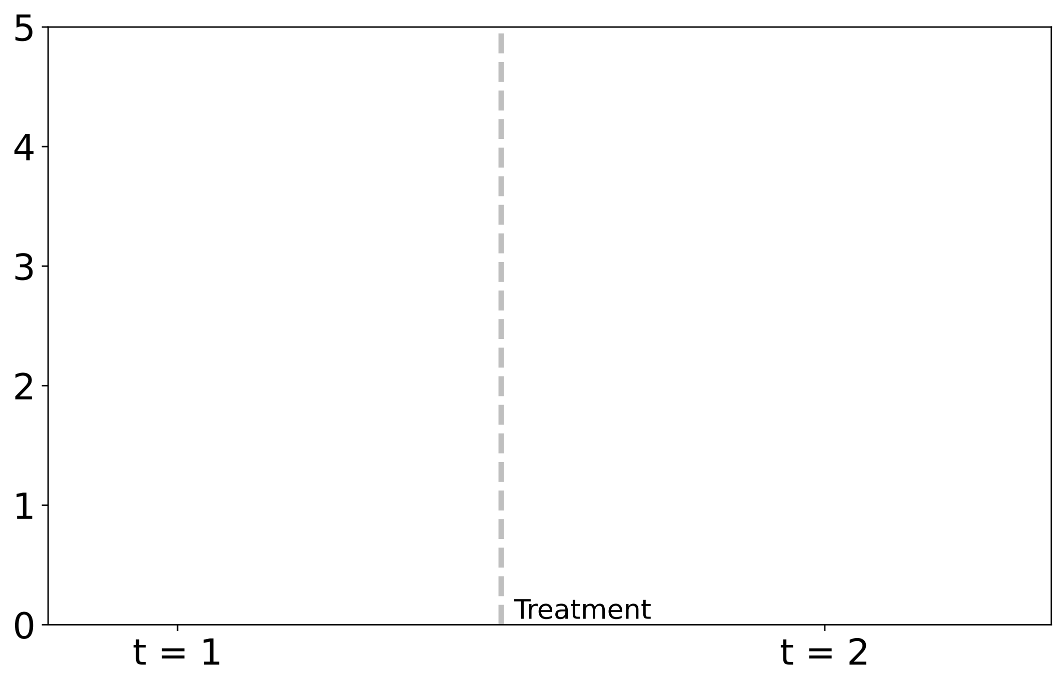
- First documented example (though not formalized): John Snow showing how cholera spread through the water in London (Snow, 1855)[1]
- Modern usage introduced formally by (Ashenfelter, 1978), applied to labor economics

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

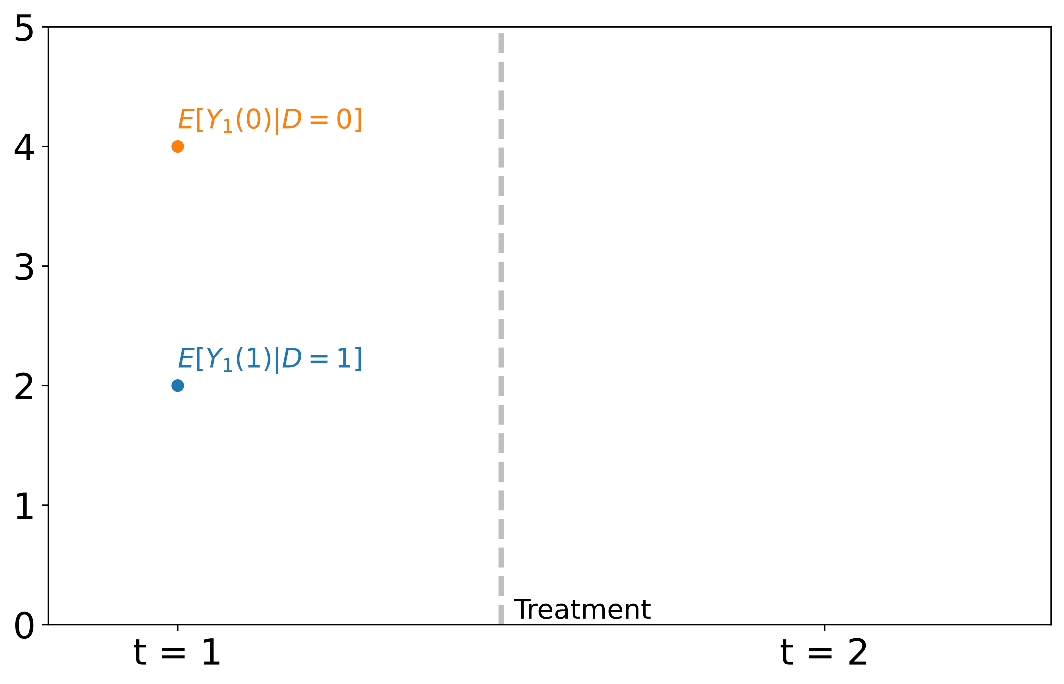
[1] Good description: https://mixtape.scunning.com/09-difference_in_differences#john-snows-cholera-hypothesis

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



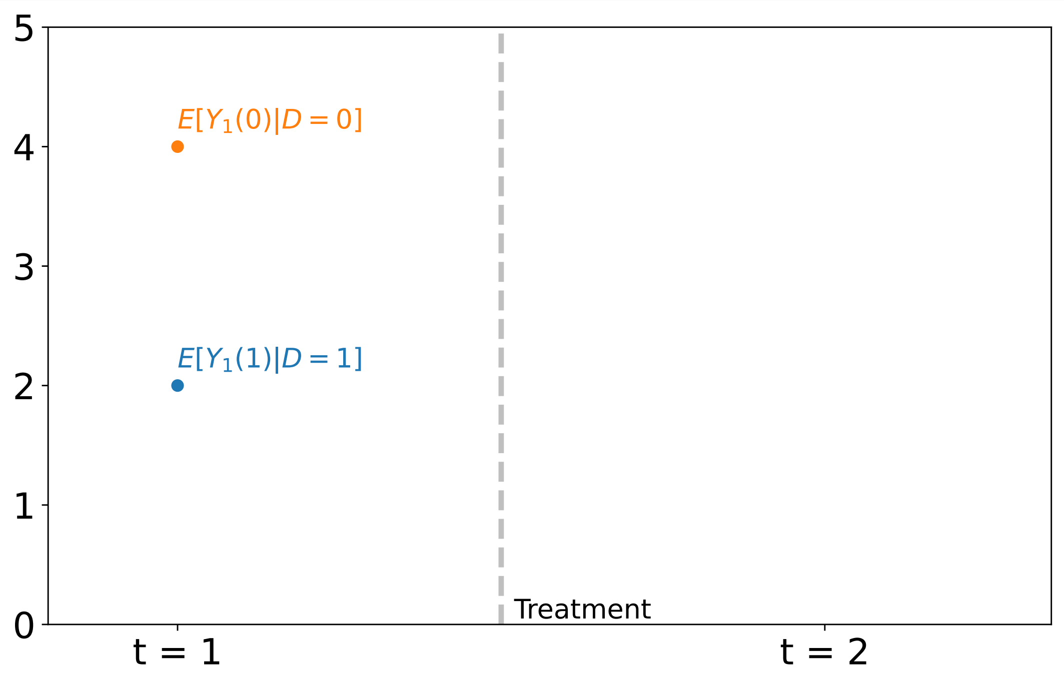
Difference-in-differences framework

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



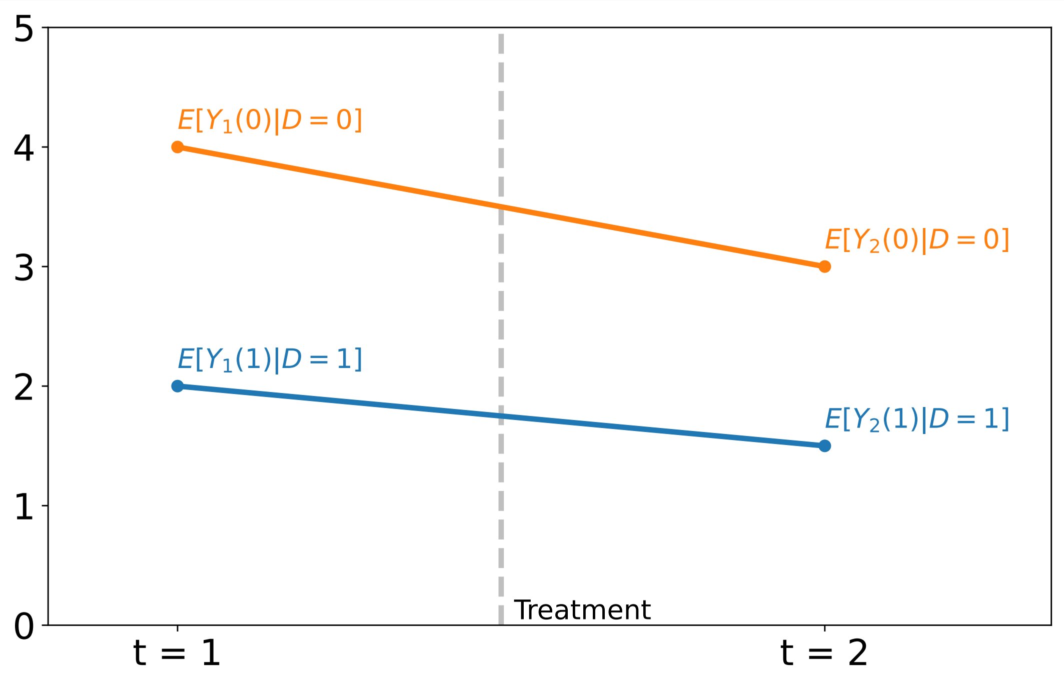
Difference-in-differences framework

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



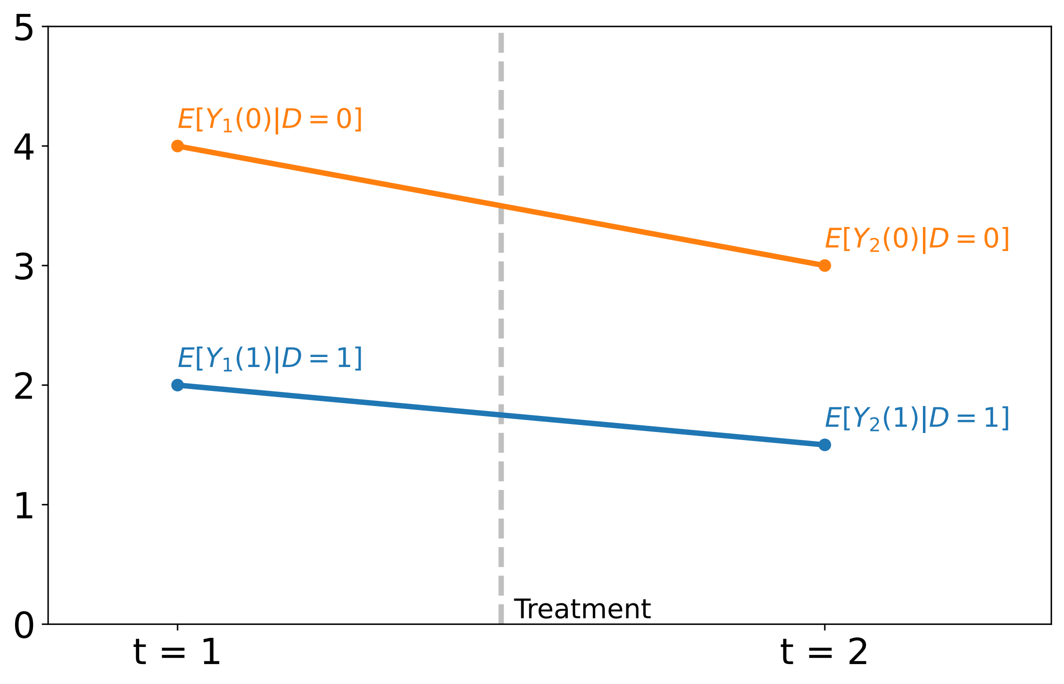
⚠️ $\mathbb{E}[Y_1(1)] = \mathbb{E}[Y_1(1)\ |D=1]\mathbb{P}(D=1) + \mathbb{E}[Y_1(1)\ |D=0]\mathbb{P}(D=0)$
but we only observe $\mathbb{E}[Y_1(1)\ |D=1]$

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



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

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

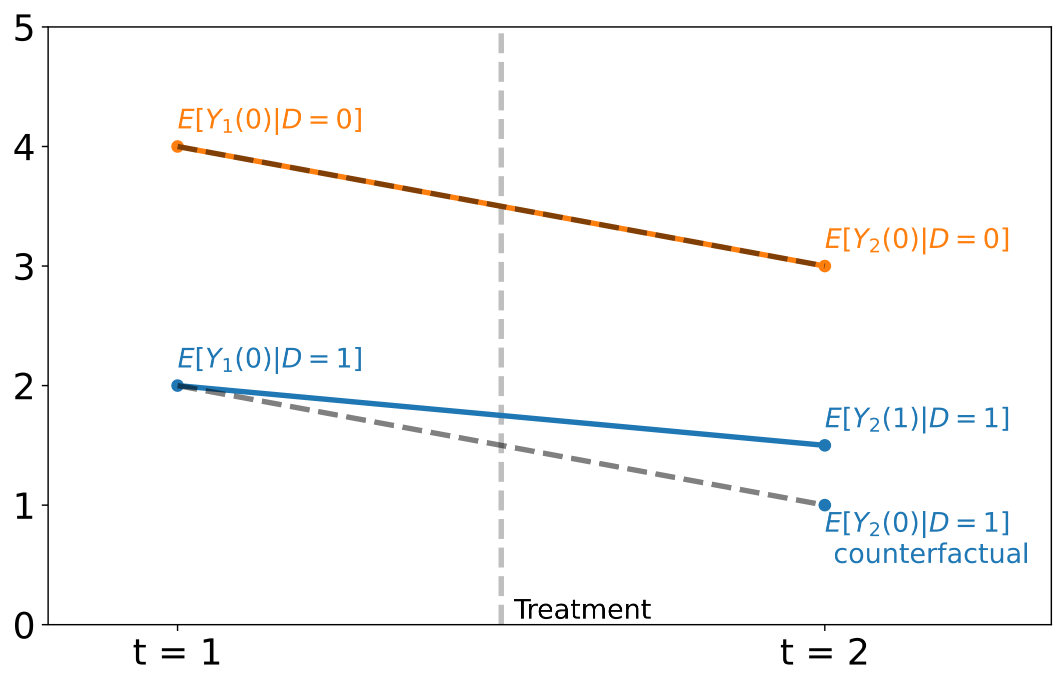


$$\tau_{\text{ATT}} = \mathbb{E}[Y_2(1)|\ D=1] - \underbrace{\mathbb{E}[Y_2(0)|\ D=1]}_{\text{unobserved since counterfactual}}$$

Difference-in-differences framework

First assumption: Parallel trends

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

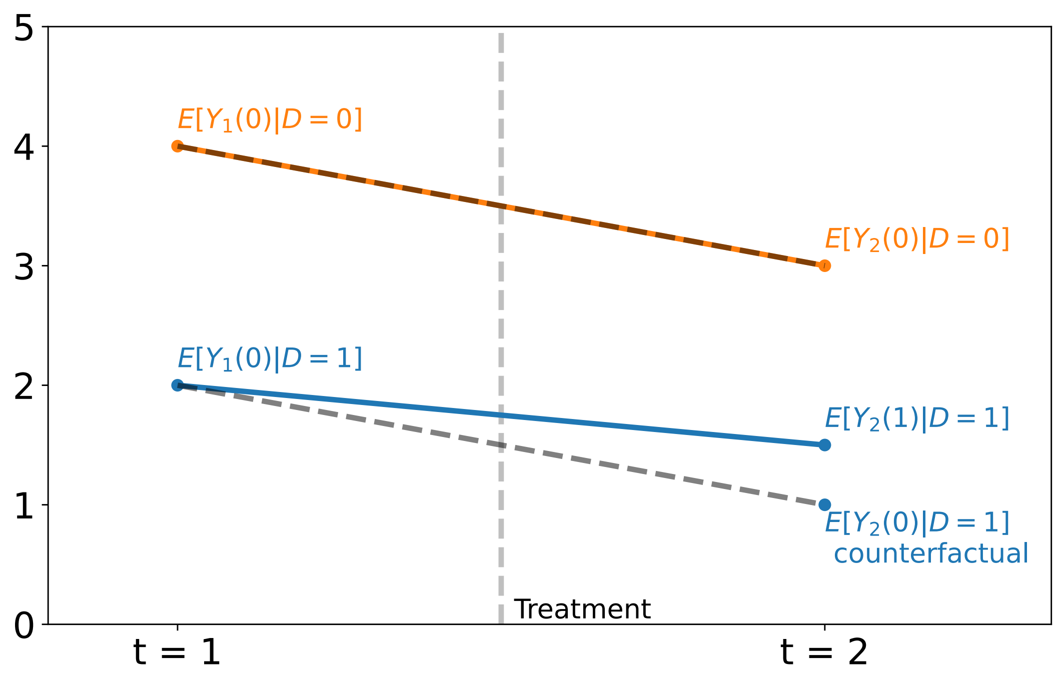


First assumption: Parallel trends[2]

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

[2] ⚠️ Strong assumption ! We will come back to it later.

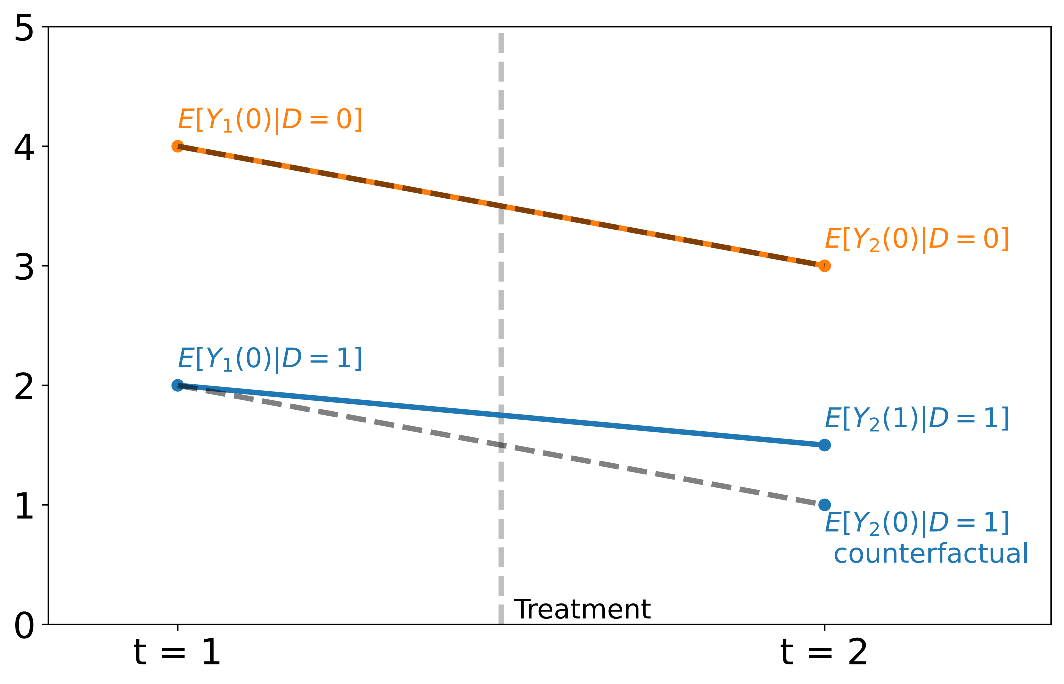
Difference-in-differences framework



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

Difference-in-differences framework

Difference-in-differences: formalization

Target effect: 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

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

Temporal consistency assumption

No anticipation of the treatment (in practice, not always true)

Assumption

$$Y_{it}(0) = Y_{it}(1) \forall t = 1, ..., H.$$

Main and strong assumption of the DID method

Assumption

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

Parallel trend assumption

Under the temporal consistency and the parallel trend assumptions, the DID estimator is unbiased ie. $\mathbb{E}[\widehat{\tau_{\text{DID}} - \tau}_{\text{SATT}}] = 0$

See (Wager, 2024) for a clear proof.

Estimation: link with two way fixed effect

In practice, DID is estimated with a two-way fixed effect model (TWFE):

$$Y_{it} \sim \alpha_i + \beta_t + A_{it}\tau \text{ where } A_{it} = D_i * (t \leq H)$$

- α_i capture the individual fixed effect
- β_t capture the time fixed effect (under pararell trend)

This link can be seen with the parallel trend assumption: $\beta_t =$

Conditional difference-in-differences

Synthetic Controls

Synthetic Controls

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

- A method for estimating 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 for the effect of taxes on sugar-based product consumption in (Puig-Codina et al., 2021), review of usage in healthcare (Bouttell et al., 2018).

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.

Bouttell, J., Craig, P., Lewsey, J., Robinson, M., & Popham, F. (2018). Synthetic control methodology as a tool for evaluating population-level health interventions. *J Epidemiol Community Health*, *72*(8), 673–678.

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.