Machine Learning for econometrics

Event studies: Causal methods for pannel data

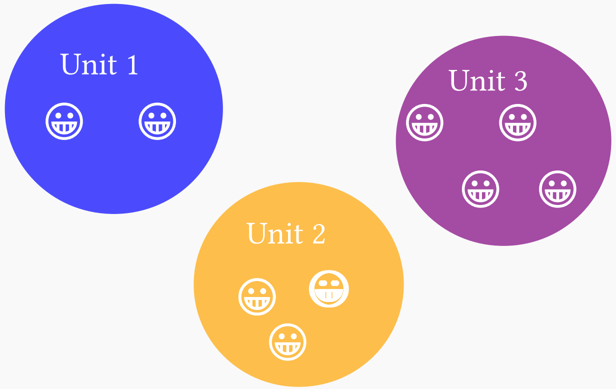
Matthieu Doutreligne

February, 11th, 2025

Motivation

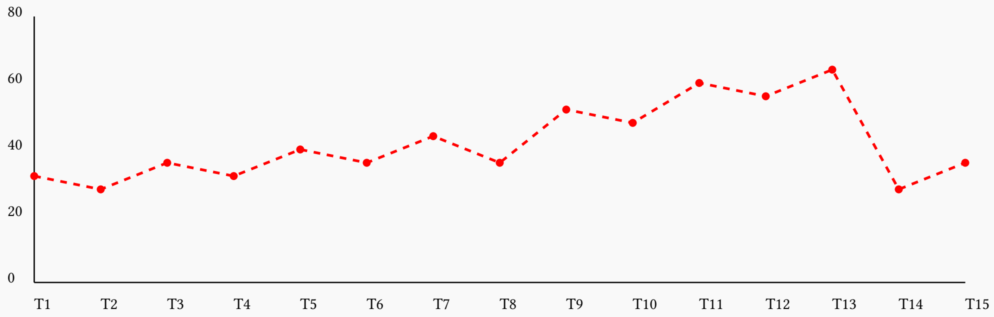
Estimation of the effect of a treatment when data is:

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



Estimation of the effect of a treatment when data is:

- Aggregated: country-level data such as employment rate, GDP…
- Longitudinal: multiple time periods (or repeated cross-sections)…



Estimation of the effect of a treatment when data is:

- Aggregated: country-level data such as employment rate, GDP…
- Longitudinal: multiple time periods (or repeated cross-sections)…
- With multiple aggregated units: countries, firms, geographical regions…

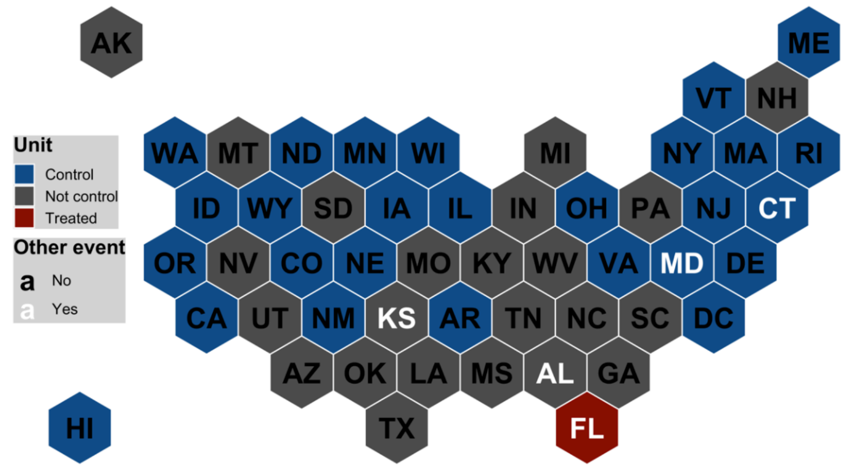
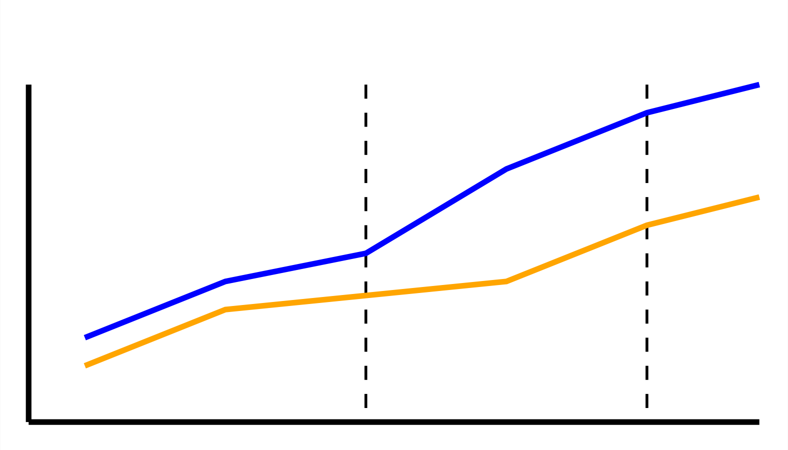


Figure from (Degli Esposti et al., 2020)

Estimation of the effect of a treatment when data is:

- Aggregated: country-level data such as employment rate, GDP…
- Longitudinal: multiple time periods (or repeated cross-sections)…
- With multiple aggregated units: countries, firms, geographical regions…
- Staggered adoption of the treatment: units adopt the policy/treatment at different times…



Setup: event studies

Estimation of the effect of a treatment when data is:

- Aggregated: country-level data such as employment rate, GDP…
- Longitudinal: multiple time periods (or repeated cross-sections)…
- With multiple aggregated units: countries, firms, geographical regions…
- Staggered adoption of the treatment: units adopt the policy/treatment at different times…

This setup is known as:

Panel data, event studies, longitudinal data, time-series data.

Examples of event studies

Archetypal questions

- Did the new marketing campaign had an effect on the sales of a product?
- Did the new tax policy had an effect on the consumption of a specific product?
- Did the guidelines on the prescription of a specific drug had an effect on the practices?

Examples of event studies

Archetypal questions

- Did the new marketing campaign had an effect on the sales of a product?
- Did the new tax policy had an effect on the consumption of a specific product?
- Did the guidelines on the prescription of a specific drug had an effect on the practices?

Modern examples

- What is the effect of the extension of Medicaid on mortality? (Miller et al., 2019)
- What is the effect of Europe's protected area policies (*Natura 2000*) on vegetation cover and on economic activity? (Grupp et al., 2023)
- Which policies achieved major carbon emission reductions? (Stechemesser et al., 2024)

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.
It should introduce *some* randomness in the treatment assignment: enforcing treatment exogeneity, ie. ignorability (ie. unconfoundedness).

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.
It should introduce *some* randomness in the treatment assignment: enforcing treatment exogeneity, ie. ignorability (ie. unconfoundedness).

Other quasi-experiment designs

- Instrumental variables: a variable that is correlated with the treatment but not with the outcome.
- Regression discontinuity design: the treatment is assigned based on a threshold of a continuous variable.

Table of contents

Reminder on difference-in-differences

History

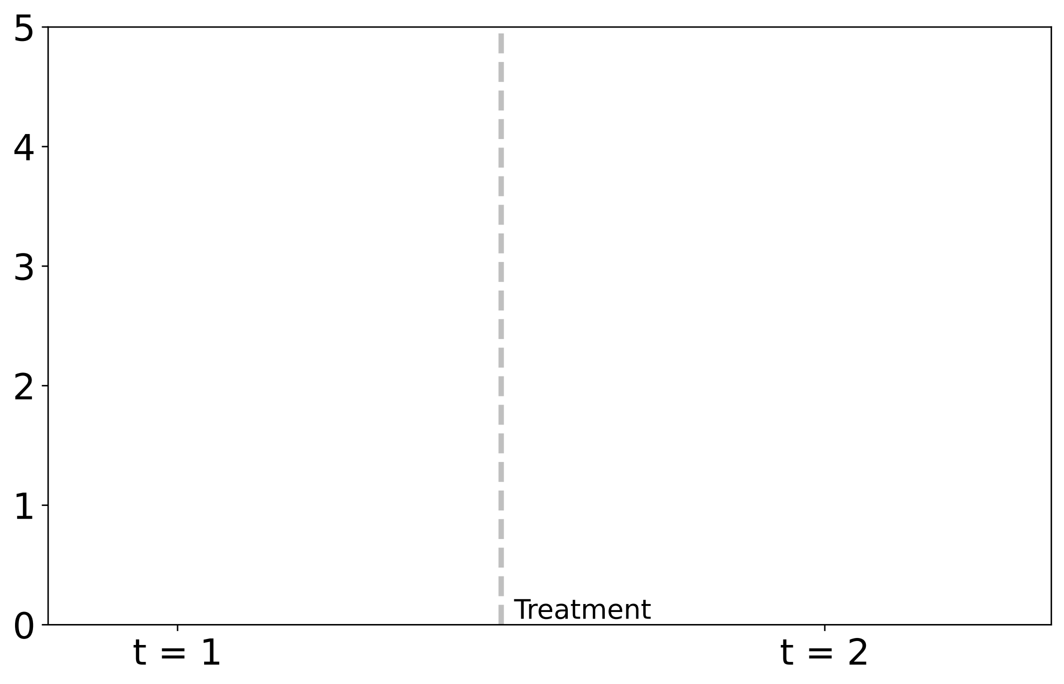
- First documented example (though not formalized): John Snow showing how cholera spread through the water in London (Snow, 1855)[1]
- Modern usage introduced formally by (Ashenfelter, 1978), applied to labor economics

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.

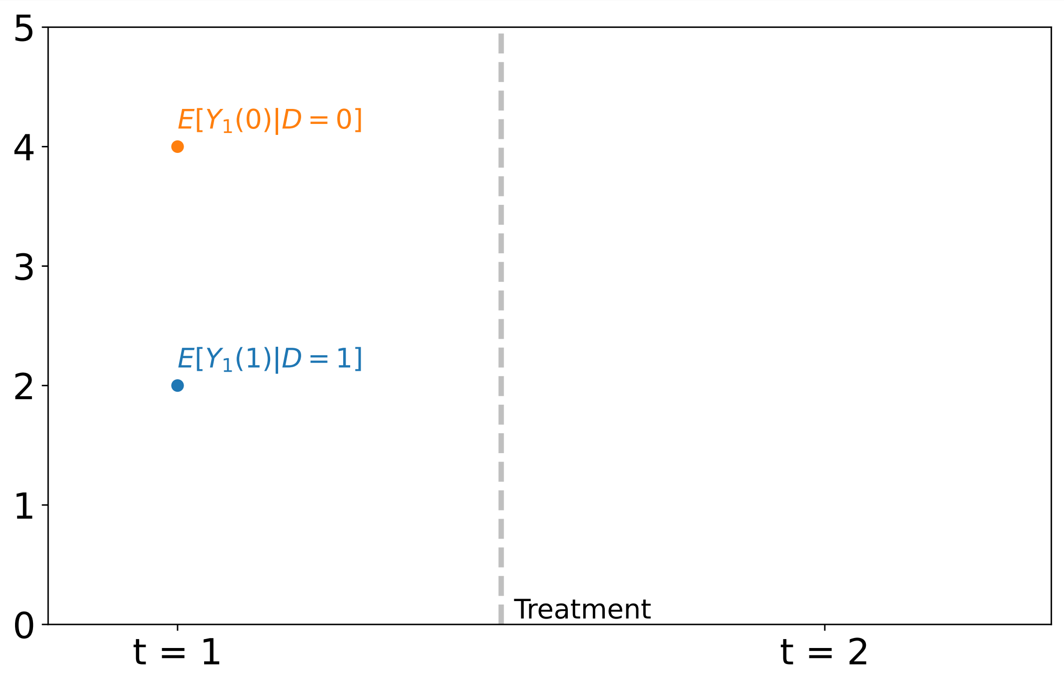
[1] Good description: https://mixtape.scunning.com/09-difference_in_differences#john-snows-cholera-hypothesis

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



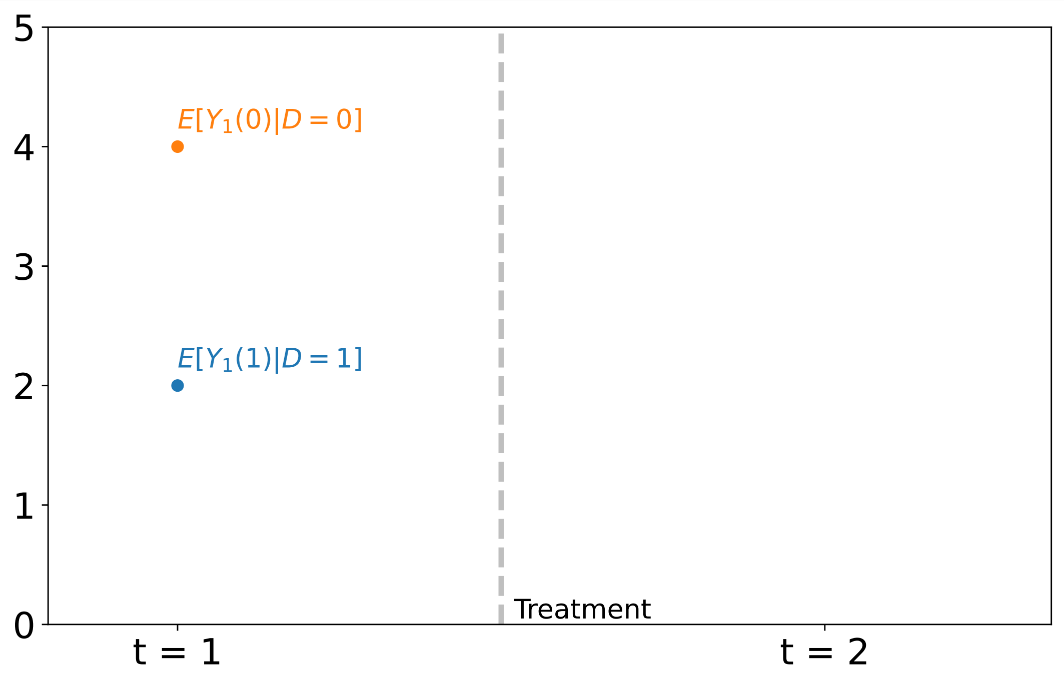
Difference-in-differences framework

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



Difference-in-differences framework

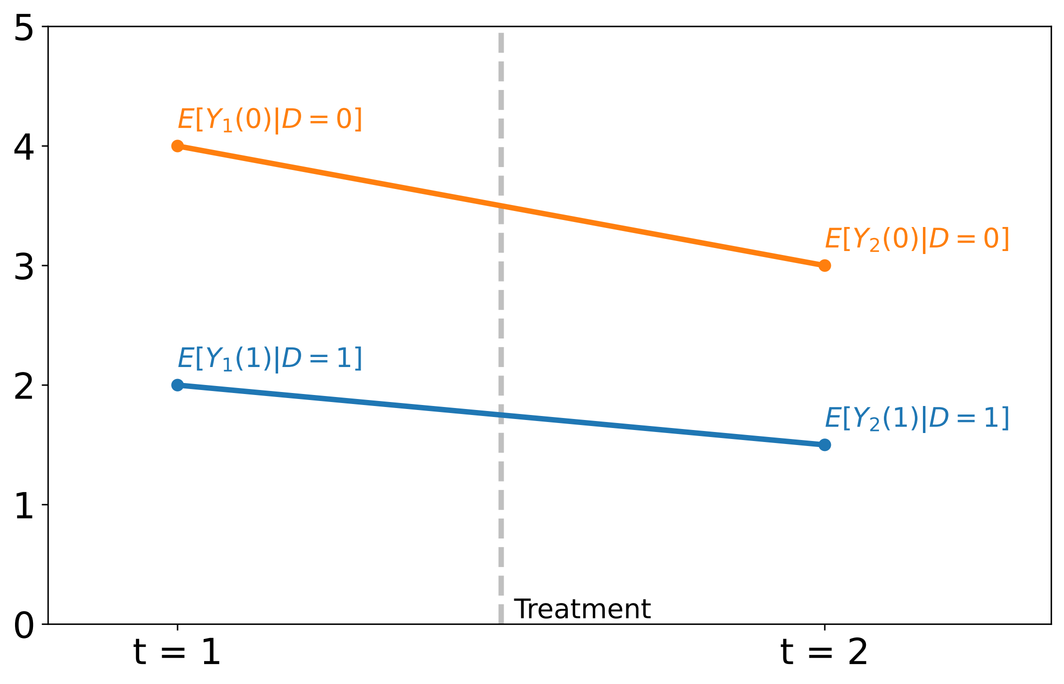
Potential outcomes: $Y_t(d)$ where $d = \{0, 1\}$ is the treatment at period 2



⚠️ $\mathbb{E}[Y_1(1)] = \underbrace{[\mathbb{E}[Y_1(1)\ |D = 0]]}_{\text{counterfactural}}\mathbb{P}(D = 0) + \underbrace{[Y_1(1)\ |D = 1]}_{\text{observed}}\mathbb{P}(D = 1)$

Difference-in-differences framework

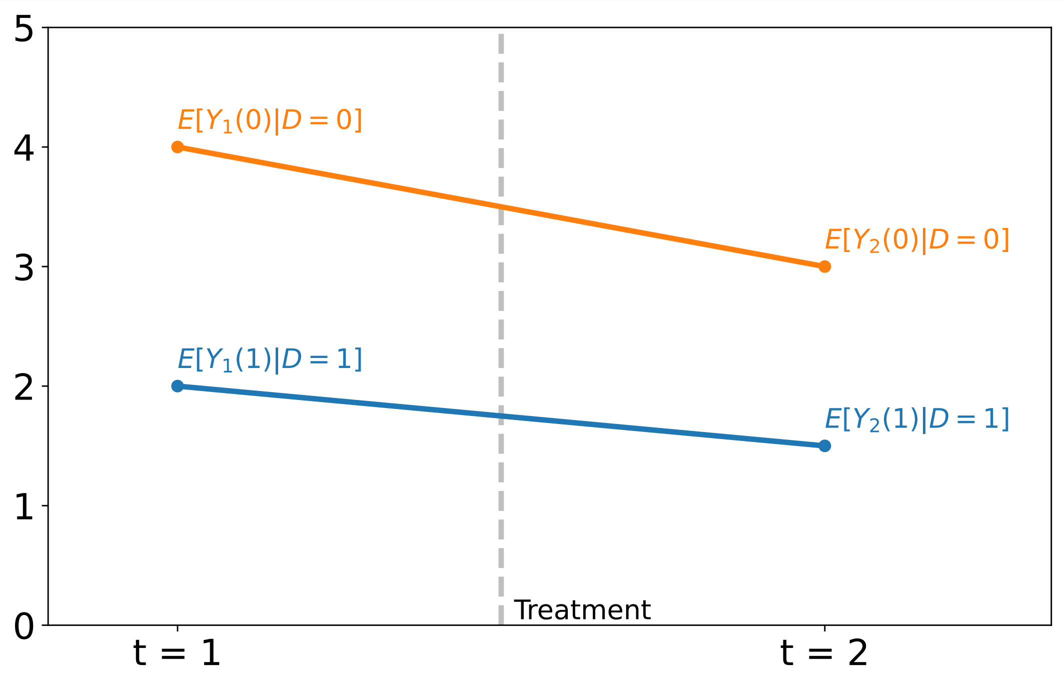
Our target is the average treatment effect on the treated (ATT)



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

Difference-in-differences framework

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

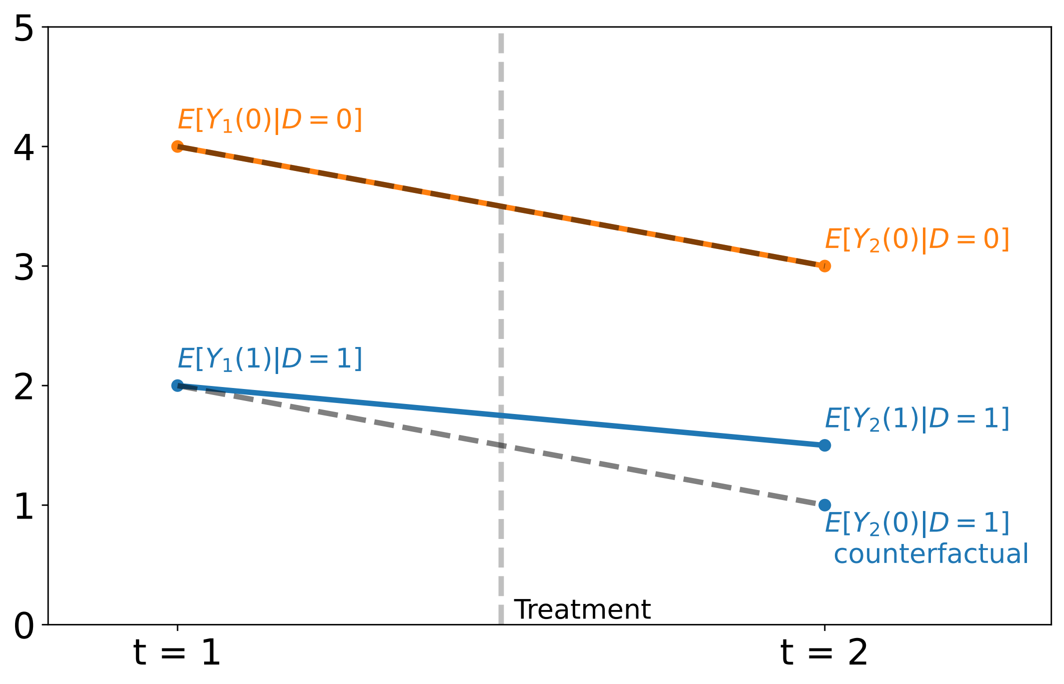


$$\tau_{\text{ATT}} = \underbrace{[Y_2(1)|\ D=1]}_{\text{treated outcome for t=2}} - \underbrace{\mathbb{E}[Y_2(0)|\ D=1]}_{\text{unobserved counterfactual}}$$

Difference-in-differences framework

First assumption, parallel trends

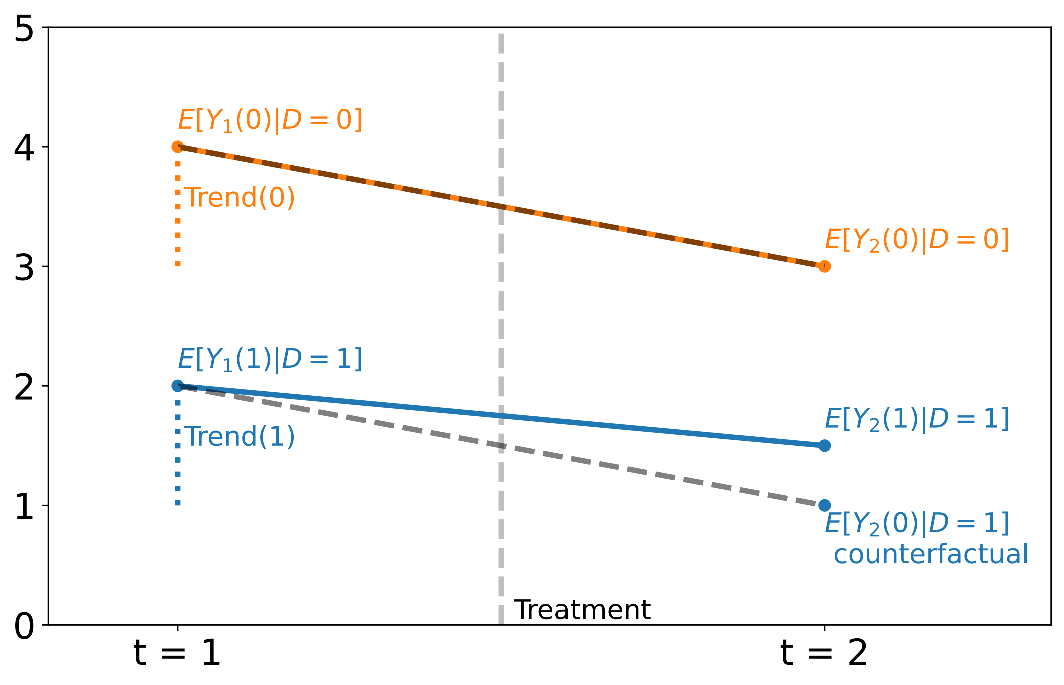
$$\mathbb{E}[Y_2(0) - Y_1(0) \mid D = 1] = \mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]$$



Difference-in-differences framework

First assumption, parallel trends

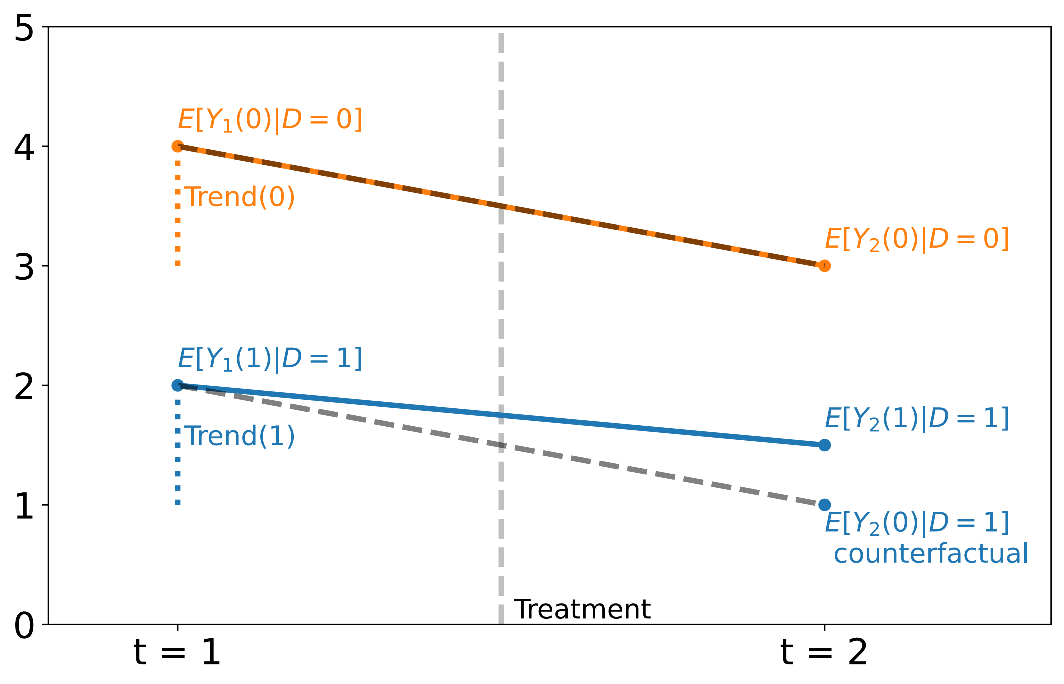
$$\underbrace{[Y_2(0) - Y_1(0) \mid D = 1]}_{\color{blue}{\text{Trend}(1)}} = \underbrace{\mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]}_{\color{orange}{\text{Trend}(0)}}$$



Difference-in-differences framework

First assumption, parallel trends

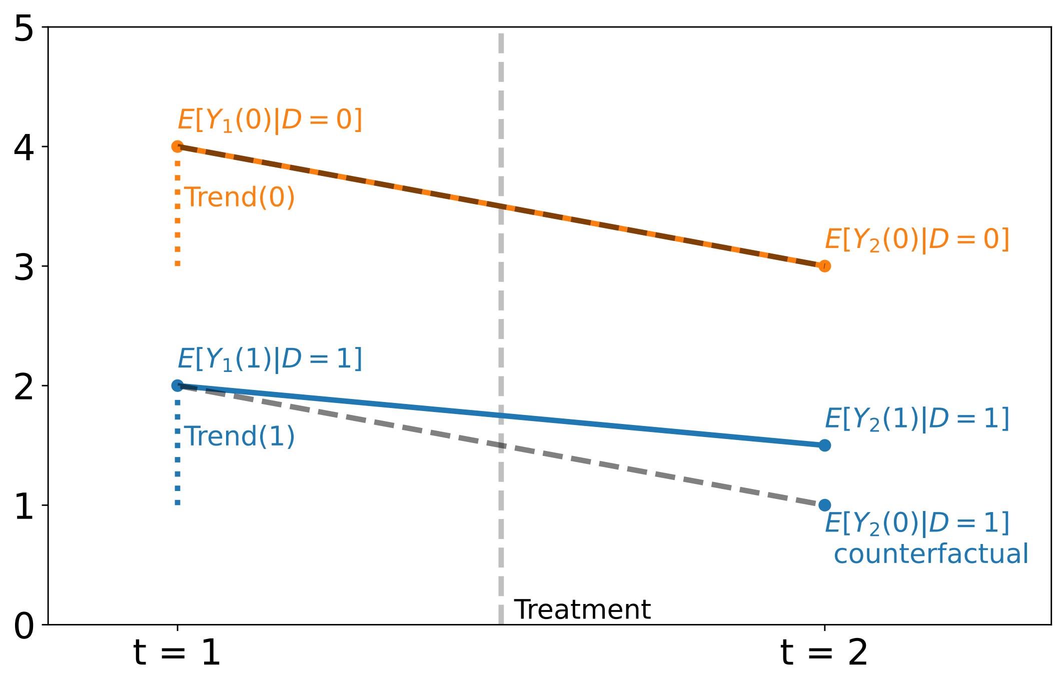
$$\mathbb{E}[Y_2(0) \mid D = 1] = \mathbb{E}[Y_1(0) \mid D = 1] + \mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]$$



Difference-in-differences framework

First assumption, parallel trends

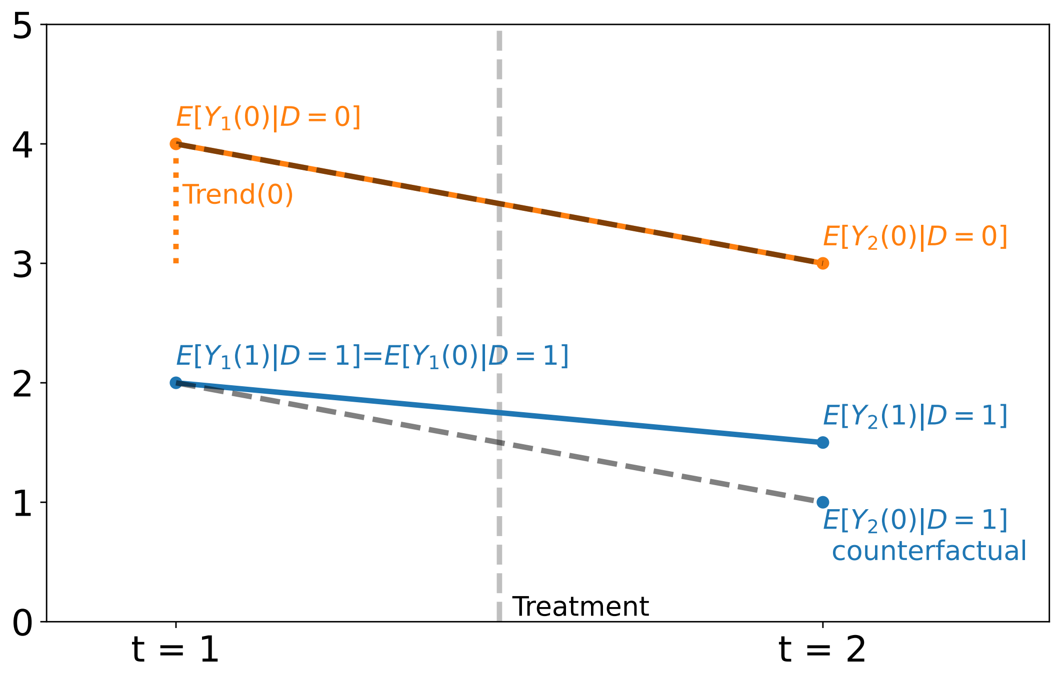
$$\mathbb{E}[Y_2(0) \mid D=1] = \underbrace{\mathbb{E}[Y_1(0) \mid D=1]}_{\text{unobserved counterfactual}} + \mathbb{E}[Y_2(0) - Y_1(0) \mid D=0]$$



Difference-in-differences framework

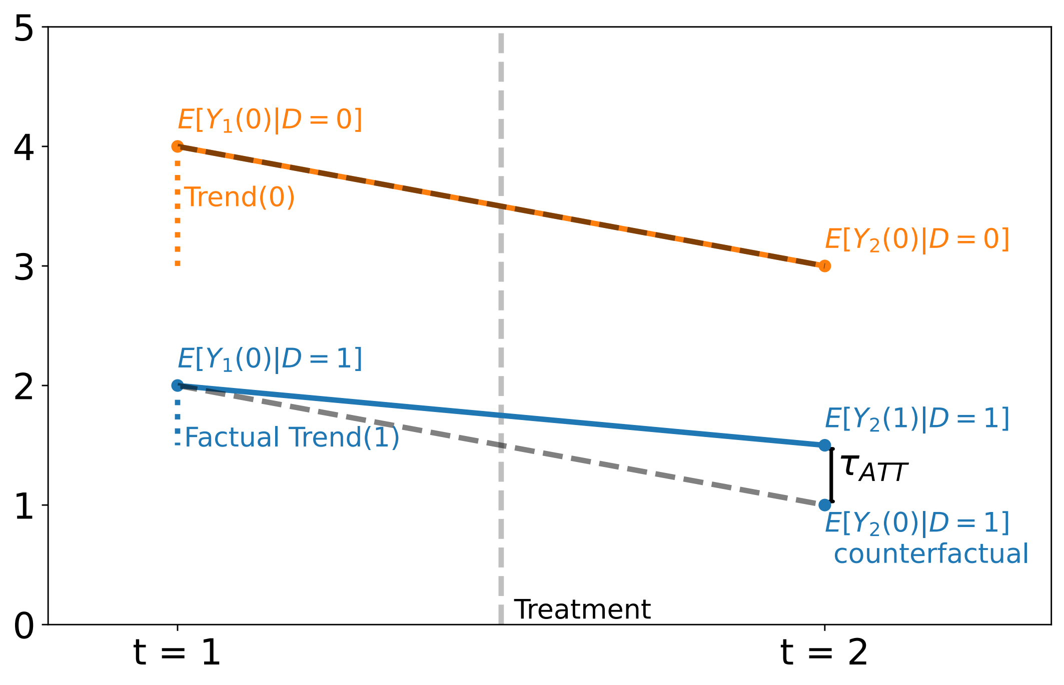
Second assumption, no anticipation of the treatment

$$\mathbb{E}[Y_1(1)|D = 1] = \mathbb{E}[Y_1(0)|D = 1]$$



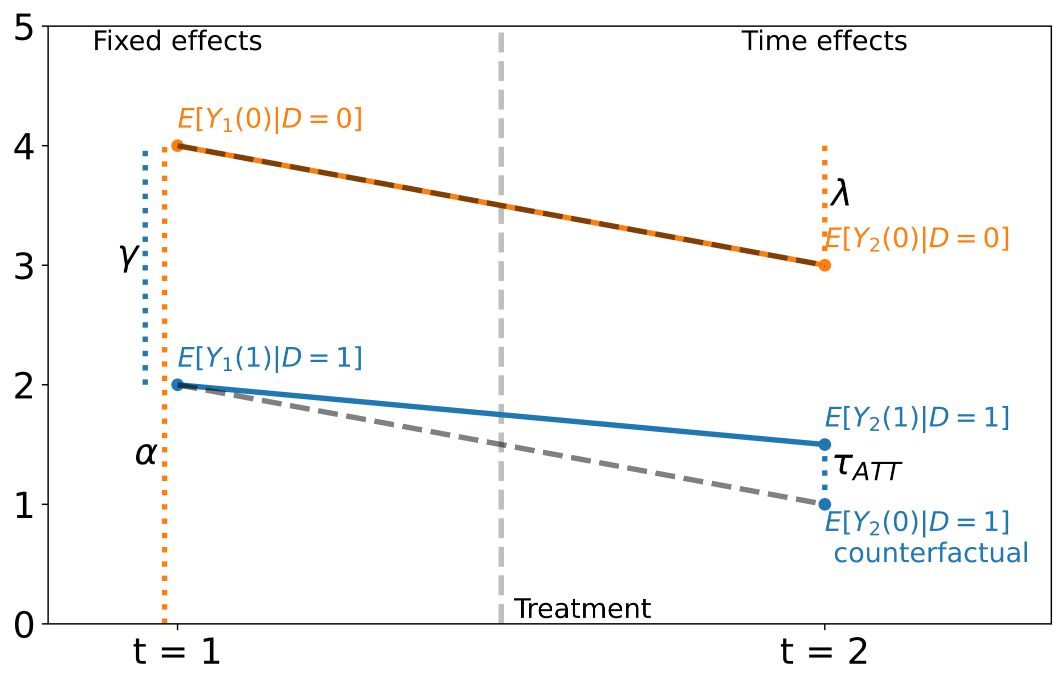
Difference-in-differences framework: identification of ATT

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



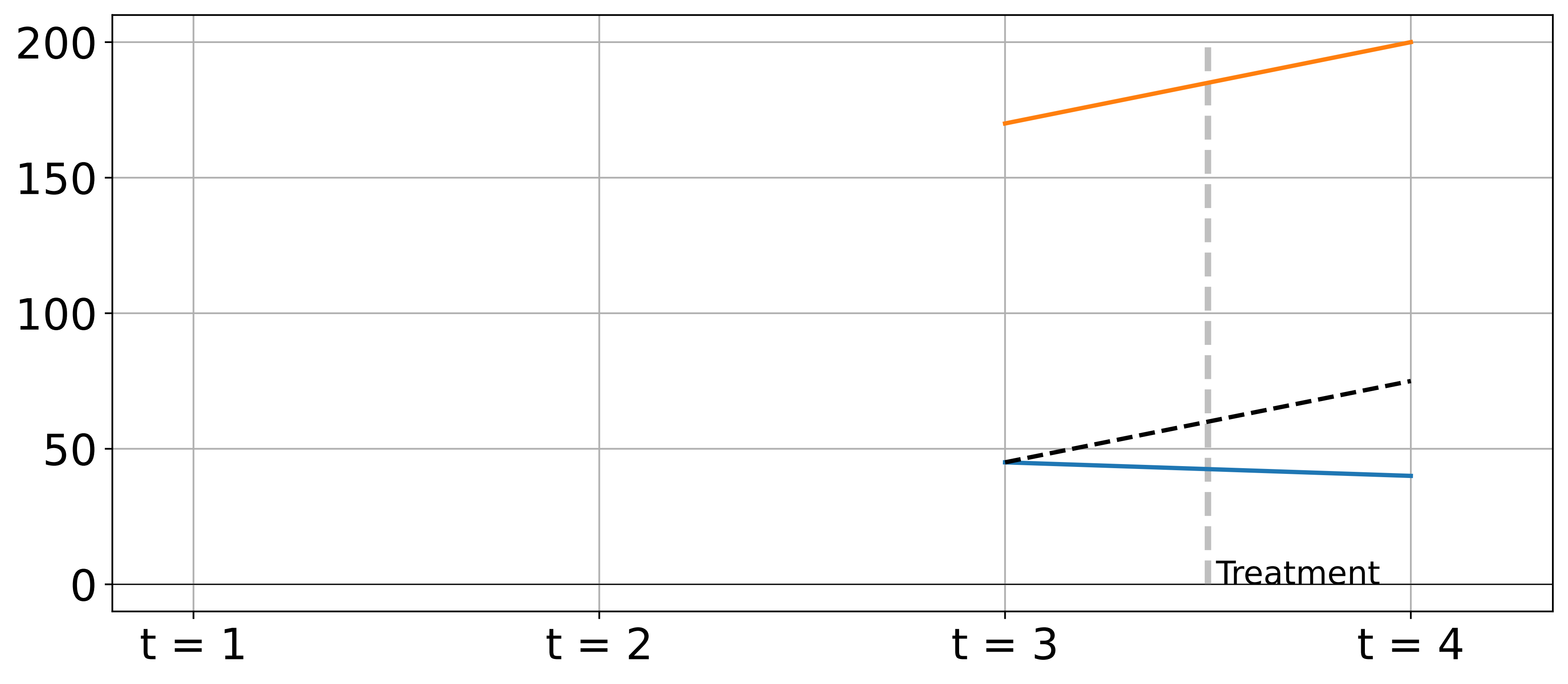
Estimation: link with two way fixed effect (TWFE)

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



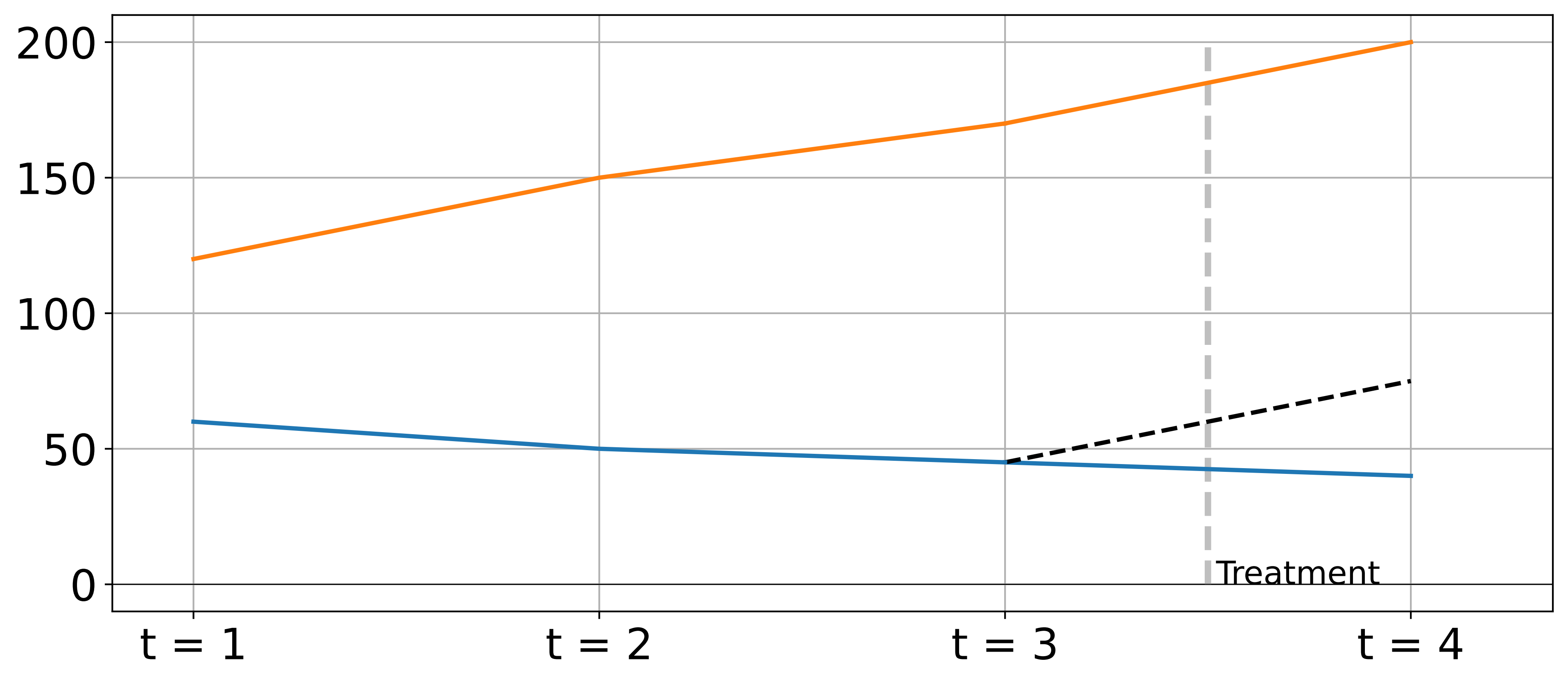
⚠️ Mechanic link: works only under parallel trends and no anticipation assumptions.

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.

Pros

- Extremely common in economics and quite simple to implement.
- Can be extended to (Wager, 2024)
  - more than two time periods: exact same formulation
  - staggered adoption of the treatment: a bit more complex

Cons

- Strong assumptions: parallel trends and no anticipation.
- Does not account for heterogeneity of treatment effect over time (De Chaisemartin & d'Haultfoeuille, 2020).

DID: Take-away

Pros

- Extremely common in economics and quite simple to implement.
- Can be extended to (Wager, 2024)
  - more than two time periods: exact same formulation
  - staggered adoption of the treatment: a bit more complex

Cons

- Strong assumptions: parallel trends and no anticipation.
- Does not account for heterogeneity of treatment effect over time (De Chaisemartin & d'Haultfoeuille, 2020).

Can we do better: ie. robust to the parallel trend assumption?

Synthetic controls

Synthetic controls

References

Introduced by (Abadie & Gardeazabal, 2003) and (Abadie et al., 2010).

Quick introduction in (Bonander et al., 2021), technical overview in (Abadie, 2021),

Synthetic controls

References

Introduced by (Abadie & Gardeazabal, 2003) and (Abadie et al., 2010).

Quick introduction in (Bonander et al., 2021), technical overview in (Abadie, 2021),

> The most important innovation in the policy evaluation literature in the last few years
>
> — (Athey & Imbens, 2017)

References

Introduced by (Abadie & Gardeazabal, 2003) and (Abadie et al., 2010).

Quick introduction in (Bonander et al., 2021), technical overview in (Abadie, 2021),

> The most important innovation in the policy evaluation literature in the last few years
>
> — (Athey & Imbens, 2017)

Idea

Find a weighted average of controls that predicts well the treated unit outcome before treatment.

Example

What is the effect of tobacco tax on cigarettes sales? (Abadie et al., 2010)

Examples of application of synthetic controls to epidemiology

- Literature review of the usage of SCM in healthcare (up to 2016): (Bouttell et al., 2018)

Some use cases

- What is the effect of UK pay-for-performance program in primary care on mortality? (Ryan et al., 2016)
- What is the effect of soda taxes on sugar-based product consumption? (Puig-Codina et al., 2021)
- What is the effect of Ohio vaccine lottery on covid-19 vaccination? (Brehm et al., 2022)
- What is the effect of wildfire storm on respiratory hospitalizations? (Sheridan et al., 2022)

Synthetic control example: California's Proposition 99 (Abadie et al., 2010)

Context

1988: 25-cent tax per pack of cigarettes, ban of on cigarette vending machines in public areas accessible by juveniles, and a ban on the individual sale of single cigarettes.

Synthetic control example: California's Proposition 99 (Abadie et al., 2010)

Context

1988: 25-cent tax per pack of cigarettes, ban of on cigarette vending machines in public areas accessible by juveniles, and a ban on the individual sale of single cigarettes.

Setup

Outcome, $Y_{j,t}$: cigarette sales per capita

Synthetic control example: California's Proposition 99 (Abadie et al., 2010)

Context

1988: 25-cent tax per pack of cigarettes, ban of on cigarette vending machines in public areas accessible by juveniles, and a ban on the individual sale of single cigarettes.

Setup

Outcome, $Y_{j,t}$: cigarette sales per capita

Treated unit, $j = 1$: California as from 1988

Synthetic control example: California's Proposition 99 (Abadie et al., 2010)

Context

1988: 25-cent tax per pack of cigarettes, ban of on cigarette vending machines in public areas accessible by juveniles, and a ban on the individual sale of single cigarettes.

Setup

Outcome, $Y_{j,t}$: cigarette sales per capita

Treated unit, $j = 1$: California as from 1988

Control units, $j \in \{2, ..J\}$: 39 other US states without similar policies

Synthetic control example: California's Proposition 99 (Abadie et al., 2010)

Context

1988: 25-cent tax per pack of cigarettes, ban of on cigarette vending machines in public areas accessible by juveniles, and a ban on the individual sale of single cigarettes.

Setup

Outcome, $Y_{j,t}$: cigarette sales per capita

Treated unit, $j = 1$: California as from 1988

Control units, $j \in \{2, ..J\}$: 39 other US states without similar policies

Time period: $t \in \{1, ..T\} = \{1970, ..2000\}$ and treatment time $T_0 = 1988$

Synthetic control example: California's Proposition 99 (Abadie et al., 2010)

Context

1988: 25-cent tax per pack of cigarettes, ban of on cigarette vending machines in public areas accessible by juveniles, and a ban on the individual sale of single cigarettes.

Setup

Outcome, $Y_{j,t}$: cigarette sales per capita

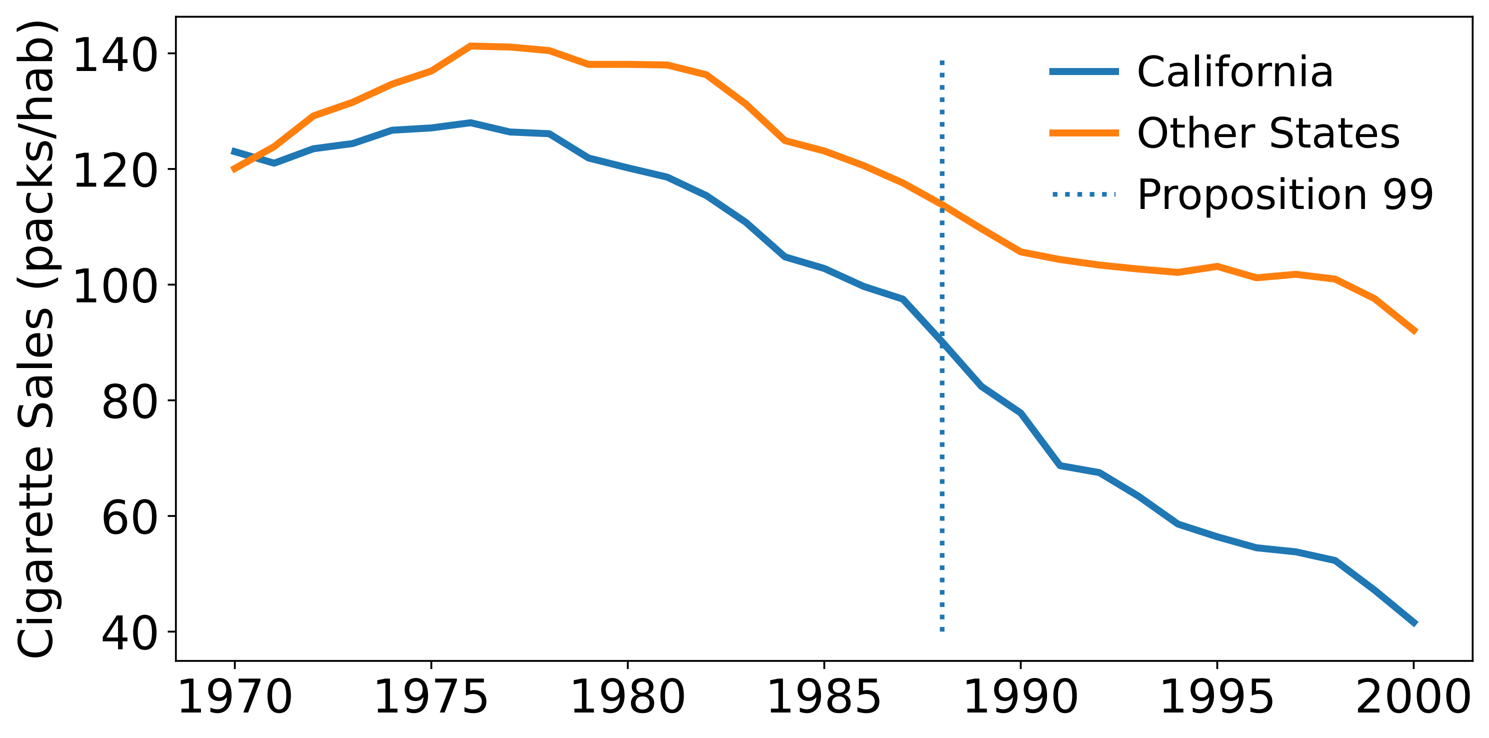
Treated unit, $j = 1$: California as from 1988

Control units, $j \in \{2, ..J\}$: 39 other US states without similar policies

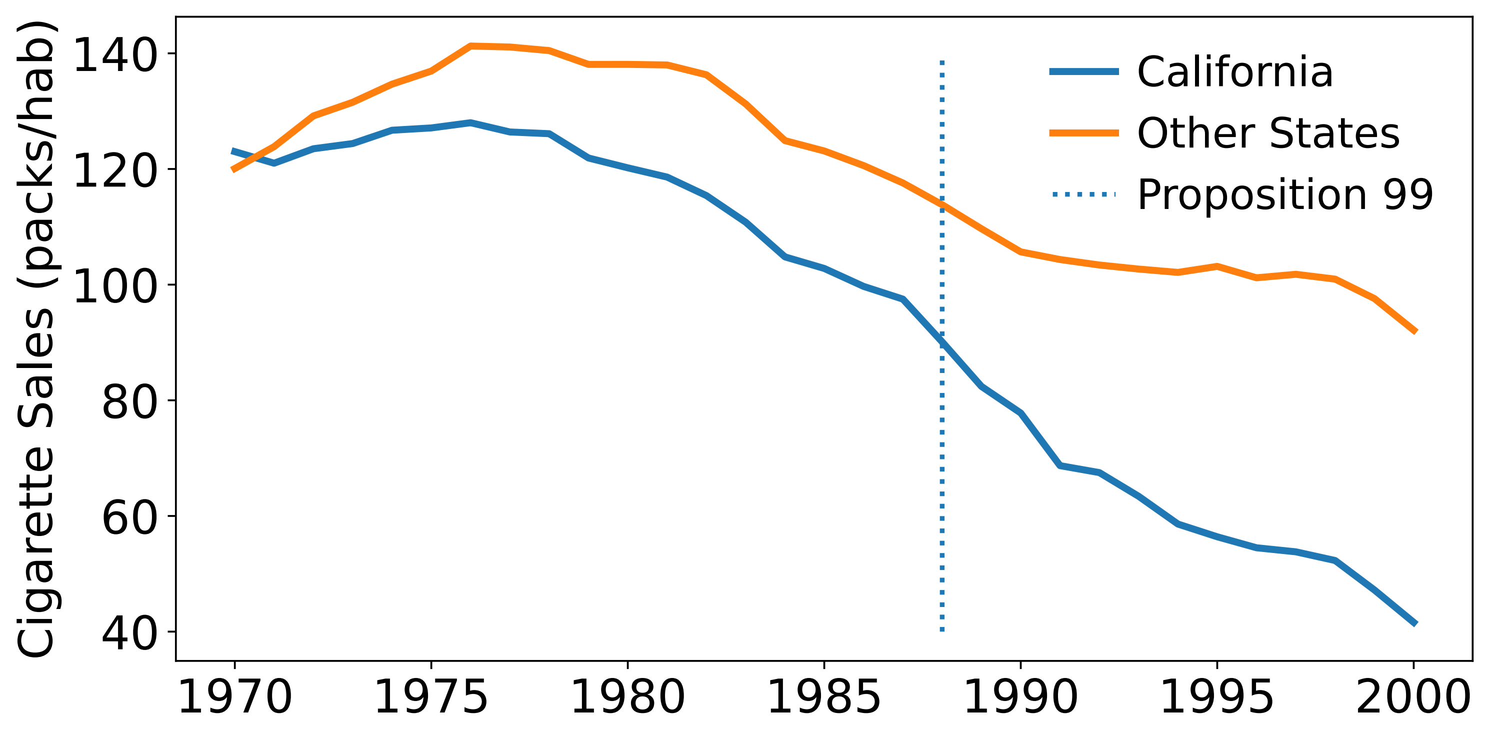
Time period: $t \in \{1, ..T\} = \{1970, ..2000\}$ **and treatment time** $T_0 = 1988$

Covariates $X_{j,t}$: cigarette price, previous cigarette sales.

Synthetic control example: plot the data

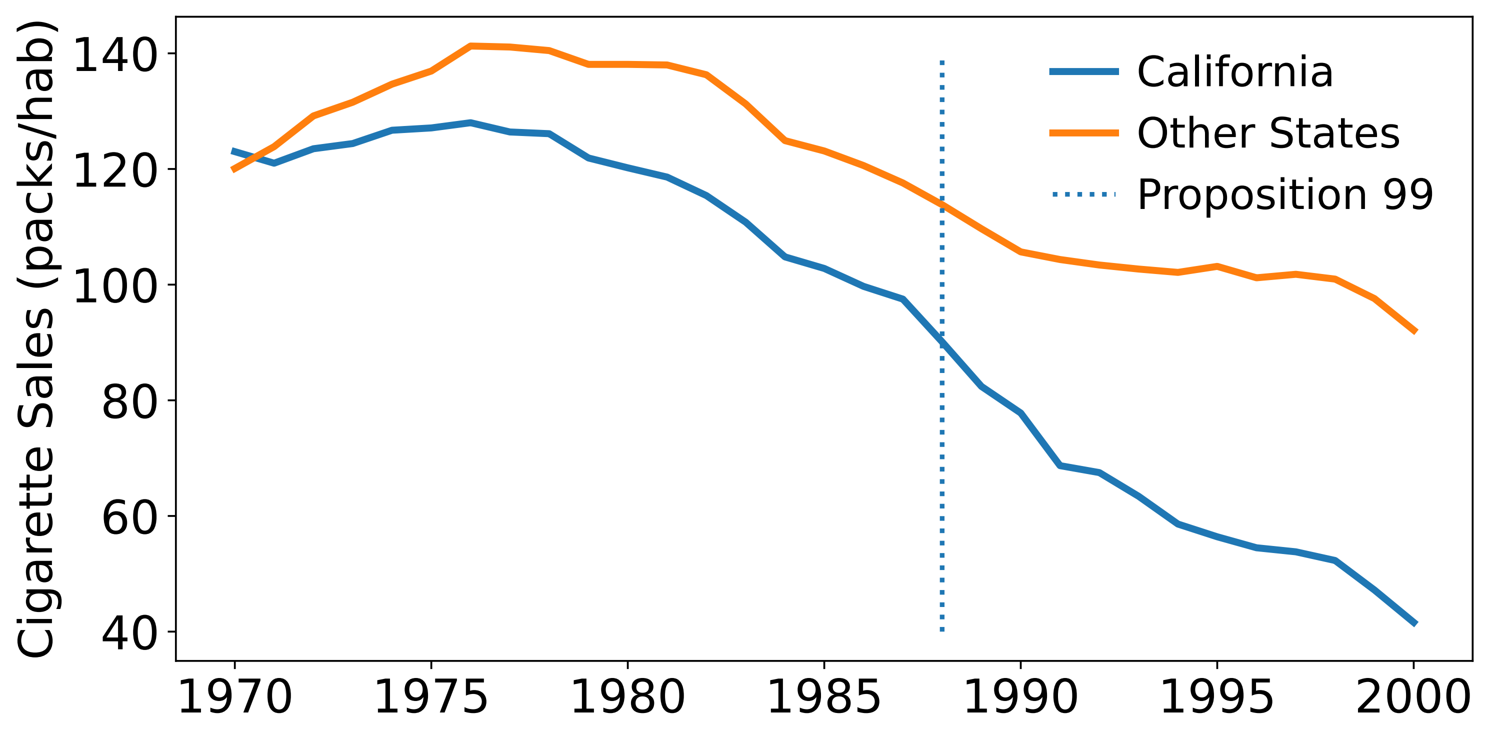


Synthetic control example: plot the data



😯 Decrease in cigarette sales in California.

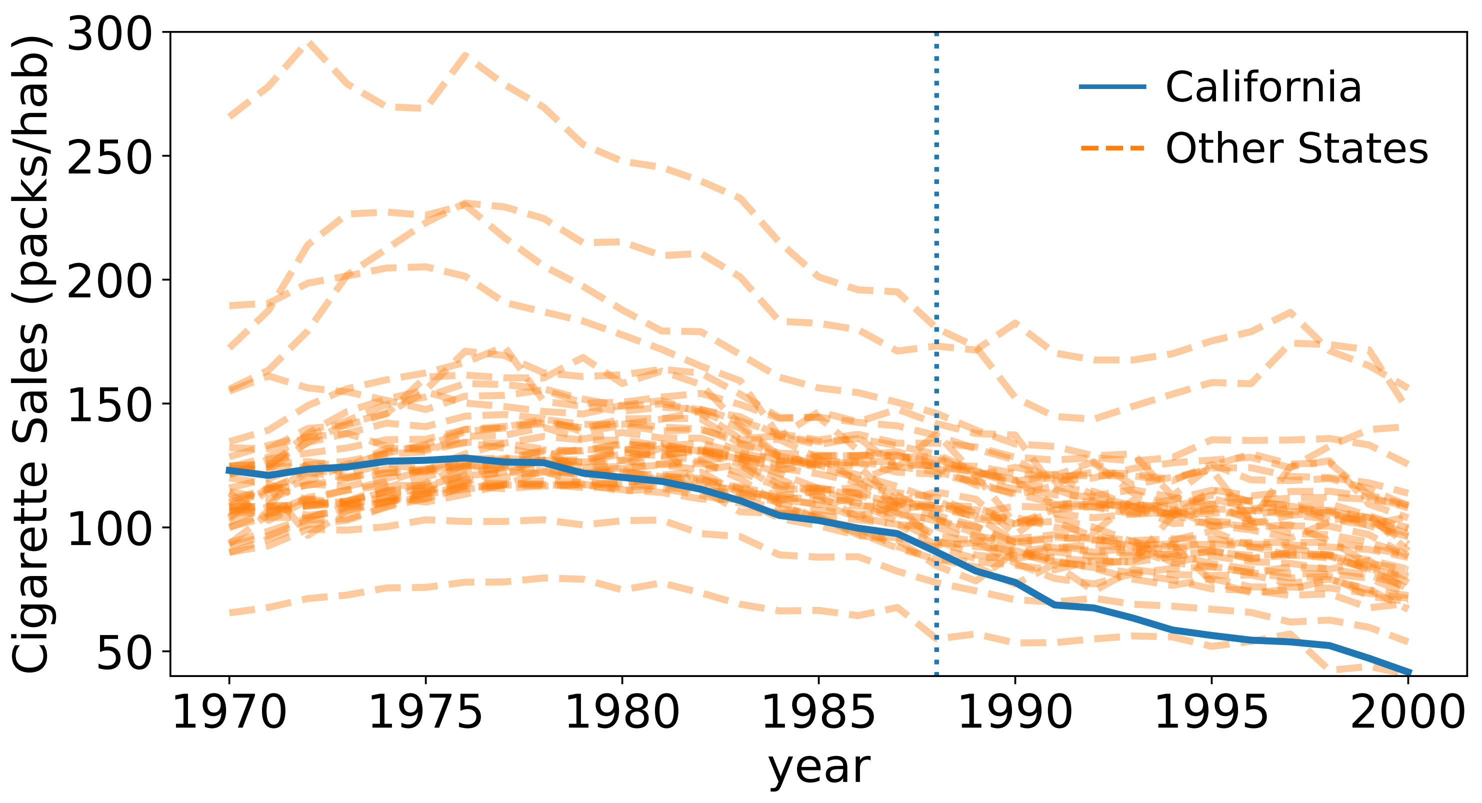
Synthetic control example: plot the data



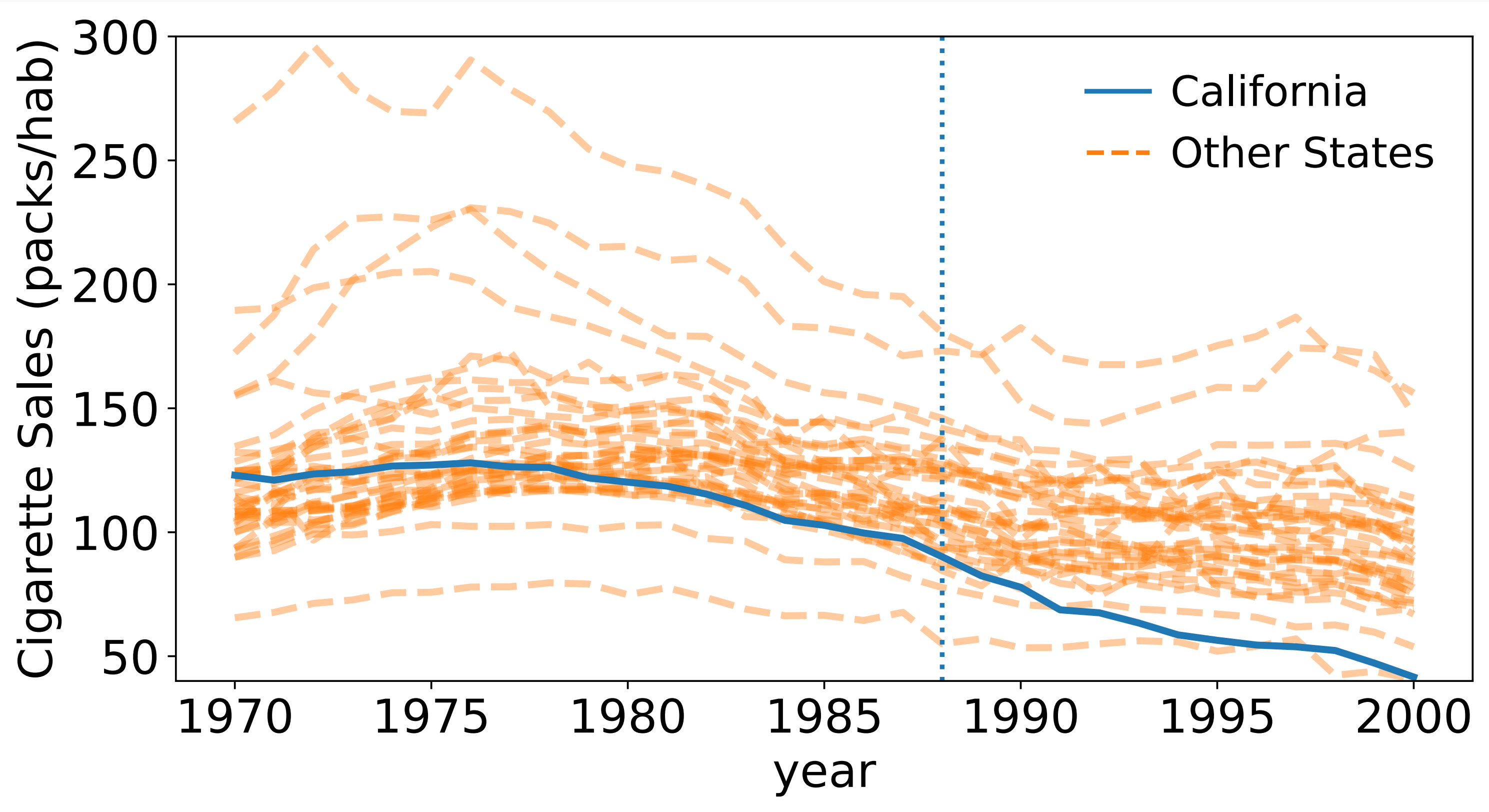
😯 Decrease in cigarette sales in California.

🤔 Decrease began before the treatment and occured also for other states.

Synthetic control example: plot the data



Synthetic control example: plot the data

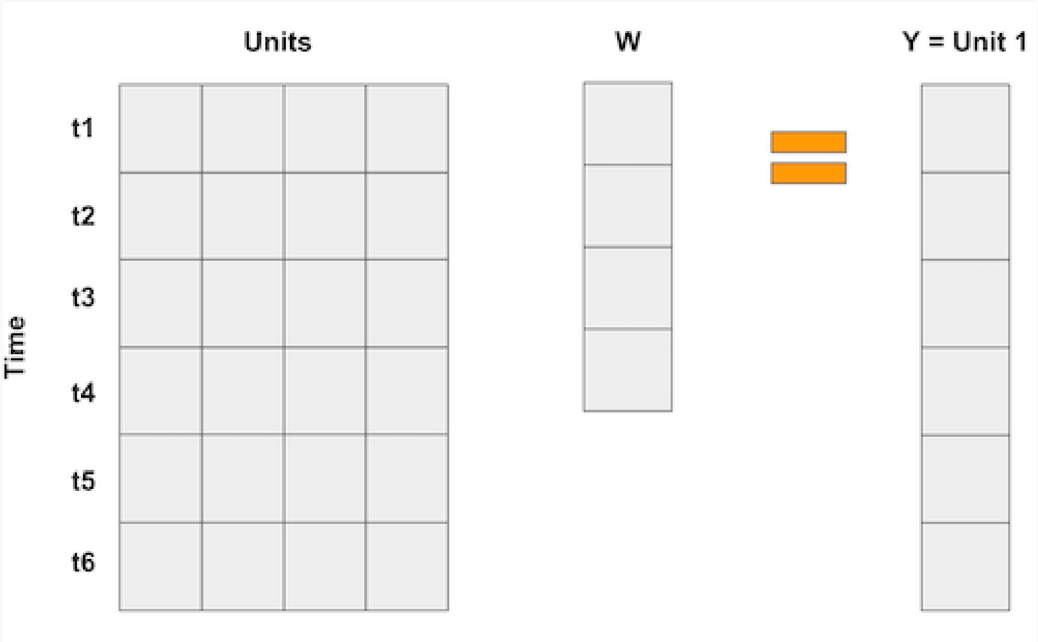


💡 Force parallel trends: Find a weighted average of other states that predicts well the pre-treatment trend of California (before $T_0 = 1988$).

Synthetic control as weighted average of control outcomes

Build a predictor for $Y_{1,t}$ (California):

$\hat{Y}_{1,t} = \sum_{j=2}^{n_0+1} \hat{w}_j Y_{j,t}$



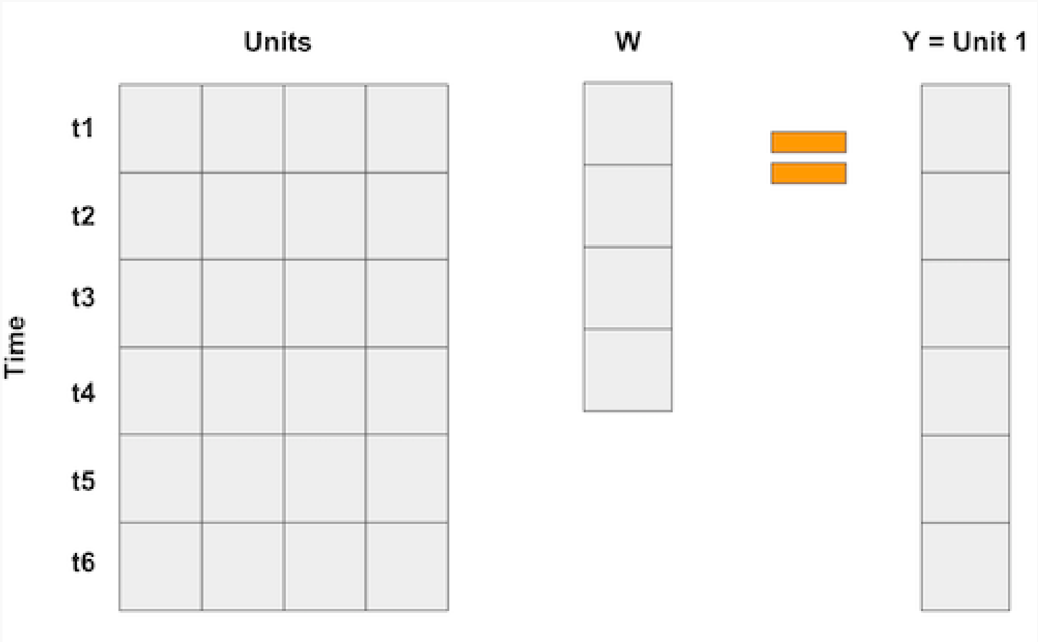
Synthetic control as weighted average of control outcomes

Build a predictor for $Y_{1,t}$ (California):

$\hat{Y}_{1,t} = \sum_{j=2}^{n_0+1} \hat{w}_j Y_{j,t}$

🤔 How to choose the weights?

Minimize some distance between the treated and the controls.



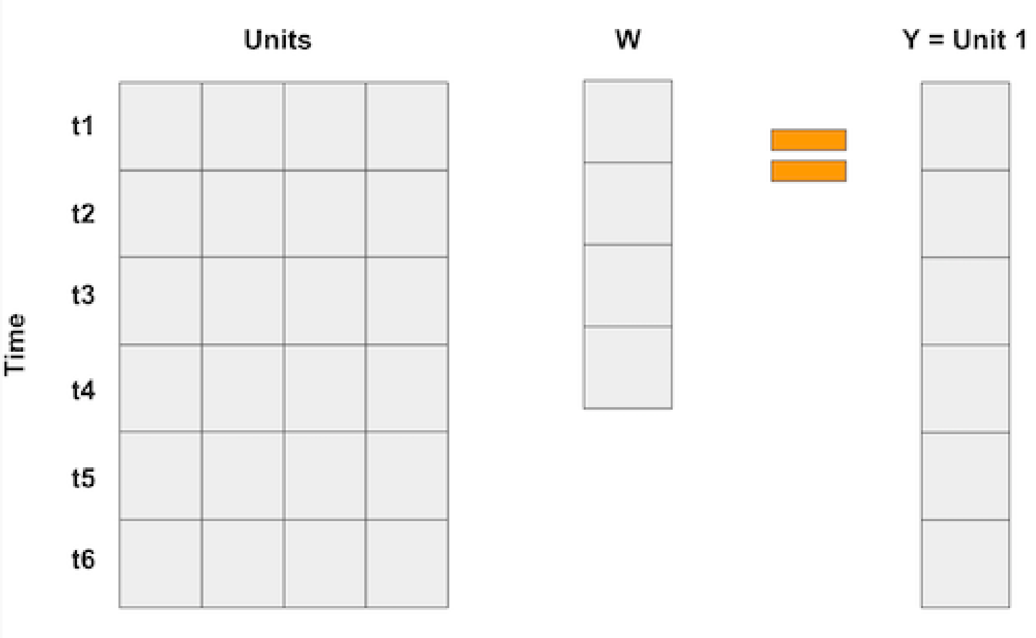
Synthetic control as weighted average of control outcomes

Build a predictor for $Y_{1,t}$ (California):

$\hat{Y}_{1,t} = \sum_{j=2}^{n_0+1} \hat{w}_j Y_{j,t}$

🤔 How to choose the weights?

Minimize some distance between the treated and the controls.



Synthetic control as weighted average of control outcomes

Build a predictor for $Y_{1,t}$ (California):

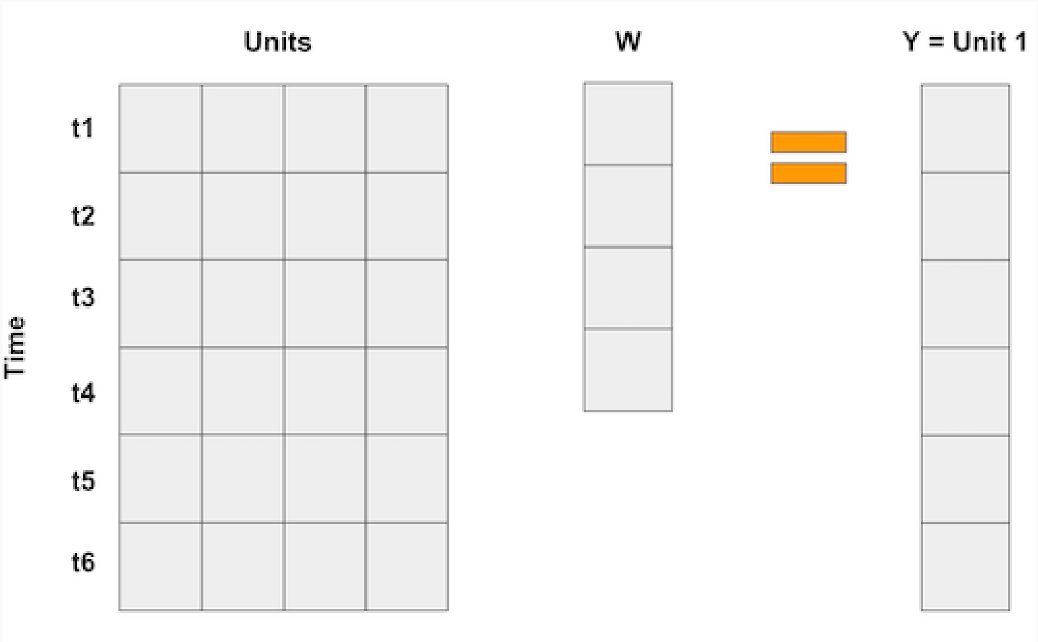
$\hat{Y}_{1,t} = \sum_{j=2}^{n_0+1} \hat{w}_j Y_{j,t}$

🤔 How to choose the weights?

Minimize some distance between the treated and the controls.

🤓 This is called a balancing estimator: kind of Inverse Probability Weighting.

Cf. (Wager, 2024, chapter 7) for details on balancing estimators.



Synthetic controls: minimization problem

Characteristics

Pre-treatment characteristics concatenate pre-treatment outcomes and other pre-treatment predictors Z_1 eg. cigarette prices:

$$X_{\text{treat}} = X_1 = \begin{pmatrix} Y_{1,1} \\ Y_{1,2} \\ .. \\ Y_{1,T_0} \\ Z_1 \end{pmatrix} \in R^{p \times 1}$$

Synthetic controls: minimization problem

Characteristics

Pre-treatment characteristics concatenate pre-treatment outcomes and other pre-treatment predictors Z_1 eg. cigarette prices:

$$X_{\text{treat}} = X_1 = \begin{pmatrix} Y_{1,1} \\ Y_{1,2} \\ .. \\ Y_{1,T_0} \\ Z_1 \end{pmatrix} \in R^{p \times 1}$$

Let the control pre-treatment characteristics be: $X_{\text{control}} = \left(X_2, .., X_{n_0+1}\right) \in R^{p \times n_0}$

Minimization problem

Synthetic controls: minimization problem

Characteristics

Pre-treatment characteristics concatenate pre-treatment outcomes and other pre-treatment predictors Z_1 eg. cigarette prices:

$$X_{\text{treat}} = X_1 = \begin{pmatrix} Y_{1,1} \\ Y_{1,2} \\ .. \\ Y_{1,T_0} \\ Z_1 \end{pmatrix} \in R^{p \times 1}$$

Let the control pre-treatment characteristics be: $X_{\text{control}} = \left(X_2, .., X_{n_0+1}\right) \in R^{p \times n_0}$

Minimization problem

$$w^* = \text{argmin}_w \; \|X_{\text{treat}} - X_{\text{control}} w\|_V^2$$

Synthetic controls: minimization problem

Characteristics

Pre-treatment characteristics concatenate pre-treatment outcomes and other pre-treatment predictors Z_1 eg. cigarette prices:

$$X_{\text{treat}} = X_1 = \begin{pmatrix} Y_{1,1} \\ Y_{1,2} \\ .. \\ Y_{1,T_0} \\ Z_1 \end{pmatrix} \in R^{p \times 1}$$

Let the control pre-treatment characteristics be: $X_{\text{control}} = \left(X_2, .., X_{n_0+1}\right) \in R^{p \times n_0}$

Minimization problem

$$w^* = \text{argmin}_w \, \|X_{\text{treat}} - X_{\text{control}} w\|_V^2$$

$$\text{where } \|X\|_V = \sqrt{X^T V X} \quad \text{with} \quad V \in \text{diag}(R^p)$$

This gives more importance to some features than others.

Synthetic controls: minimization problem

Characteristics

Pre-treatment characteristics concatenate pre-treatment outcomes and other pre-treatment predictors Z_1 eg. cigarette prices:

$$X_{\text{treat}} = X_1 = \begin{pmatrix} Y_{1,1} \\ Y_{1,2} \\ .. \\ Y_{1,T_0} \\ Z_1 \end{pmatrix} \in R^{p \times 1}$$

Let the control pre-treatment characteristics be: $X_{\text{control}} = \left(X_2, .., X_{n_0+1}\right) \in R^{p \times n_0}$

Minimization problem with constraints

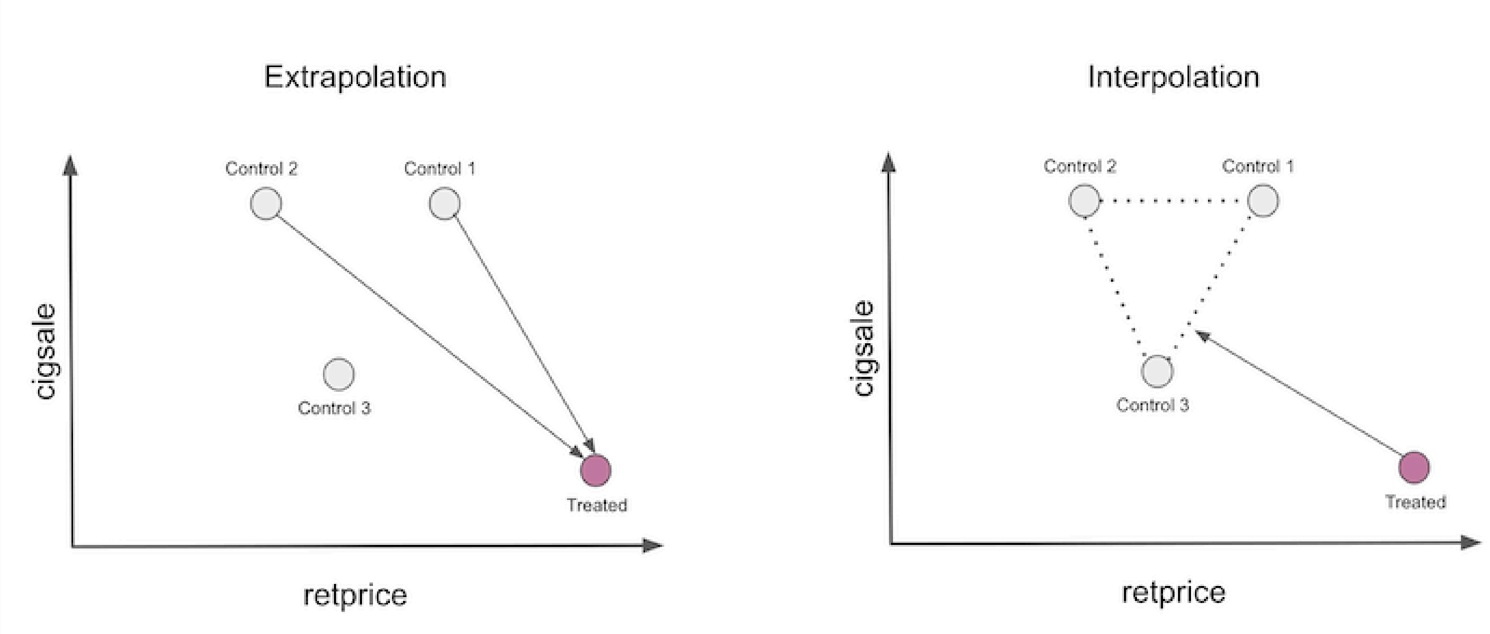
$$w^* = \text{argmin}_w \; \|X_{\text{treat}} - X_{\text{control}} w\|_V^2$$

$$s.t. \; w_j \geq 0,$$

$$\sum_{j=2}^{n_0+1} w_j = 1$$

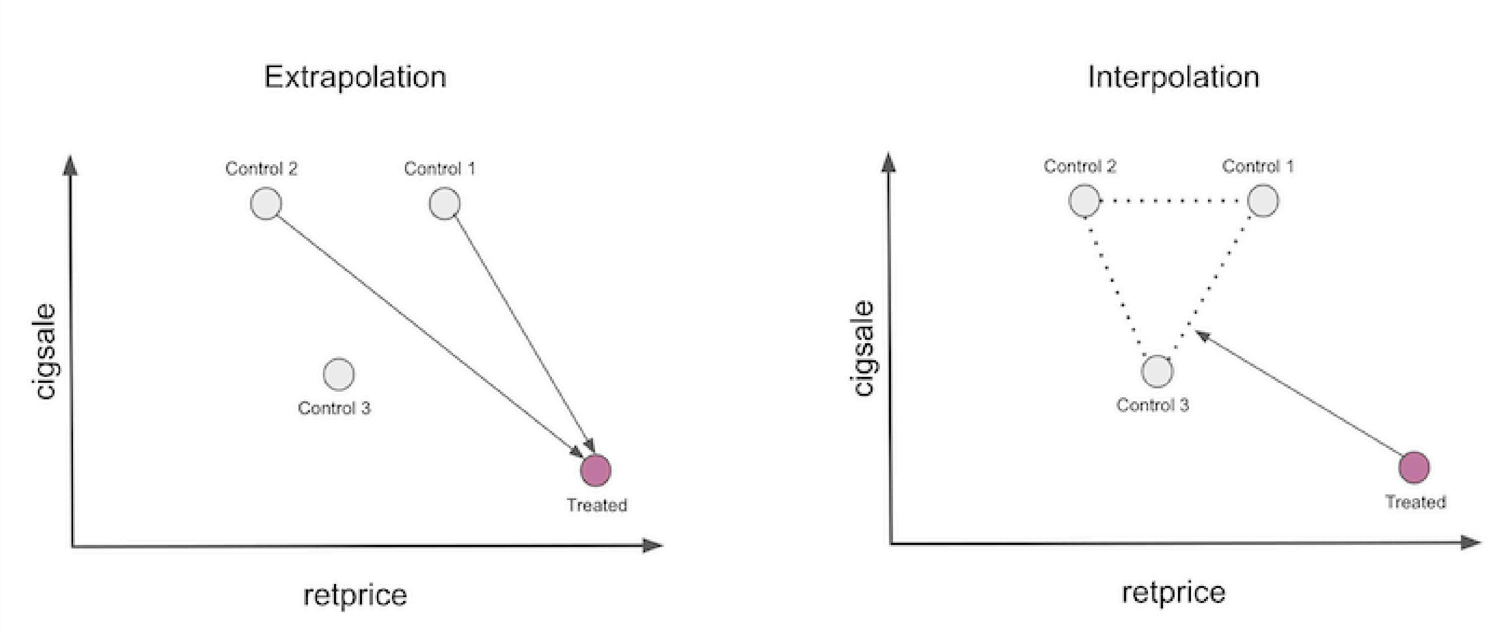
Synthetic controls: Why choose positive weights summing to one?

This is called interpolation (vs extrapolation)



Synthetic controls: Why choose positive weights summing to one?

This is called interpolation (vs extrapolation)



Interpolation enforces regularization, thus limits overfitting

Same kind of regularization than L1 norm in Lasso: forces some coefficient to be zero.

Synthetic controls: Extrapolation failure with unconstrained weight

$p = 2T_0$ covariates:

$$X_j = \begin{pmatrix} Y_{j,1} \\ .. \\ Y_{j,T_0} \\ Z_{j,1} \\ .. \\ Z_{j,T_0} \end{pmatrix}^T \in R^{2T_0}$$

Y cigarette sales, Z cigarette prices.

Synthetic controls: Extrapolation failure with unconstrained weight

$p = 2T_0$ covariates:

$$X_j = \begin{pmatrix} Y_{j,1} \\ .. \\ Y_{j,T_0} \\ Z_{j,1} \\ .. \\ Z_{j,T_0} \end{pmatrix}^T \in R^{2T_0}$$

Y cigarette sales, Z cigarette prices.

Model: $\underbrace{X_{\text{treat}}}_{p\times 1} \sim \underbrace{X_{\text{control}}}_{p\times n_0} \underbrace{w}_{n_0}$

-> simple linear regression estimated by OLS

Synthetic controls: Extrapolation failure with unconstrained weight

$p = 2T_0$ covariates:

$$X_j = \begin{pmatrix} Y_{j,1} \\ .. \\ Y_{j,T_0} \\ Z_{j,1} \\ .. \\ Z_{j,T_0} \end{pmatrix}^T \in R^{2T_0}$$

Y cigarette sales, Z cigarette prices.

Model: $\underbrace{X_{\text{treat}}}_{p\times 1} \sim \underbrace{X_{\text{control}}}_{p\times n_0} \underbrace{w}_{n_0}$

Prediction: $\hat{Y}_{\text{synth}} = \left(Y_{t,j}\right)_{\substack{t=1..T \\ j=2..n_0+1}} w$

Synthetic controls: Extrapolation failure with unconstrained weight

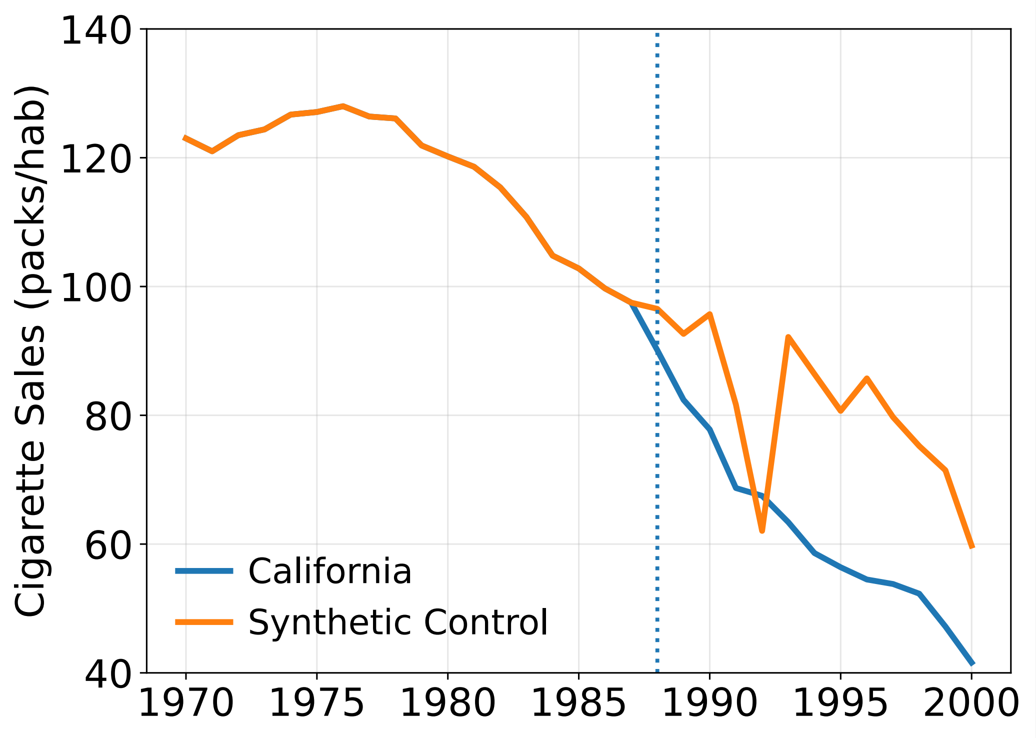
$p = 2T_0$ covariates:

$$X_j = \begin{pmatrix} Y_{j,1} \\ .. \\ Y_{j,T_0} \\ Z_{j,1} \\ .. \\ Z_{j,T_0} \end{pmatrix}^T \in R^{2T_0}$$

Y cigarette sales, Z cigarette prices.

Model: $\underbrace{X_{\text{treat}}}_{p\times 1} \sim \underbrace{X_{\text{control}}}_{p\times n_0} \underbrace{w}_{n_0}$

Prediction: $\hat{Y}_{\text{synth}} = \left(Y_{t,j}\right)_{\substack{t=1..T \\ j=2..n_0+1}} w$



Synthetic controls: Extrapolation failure with unconstrained weight

$p = 2T_0$ covariates:

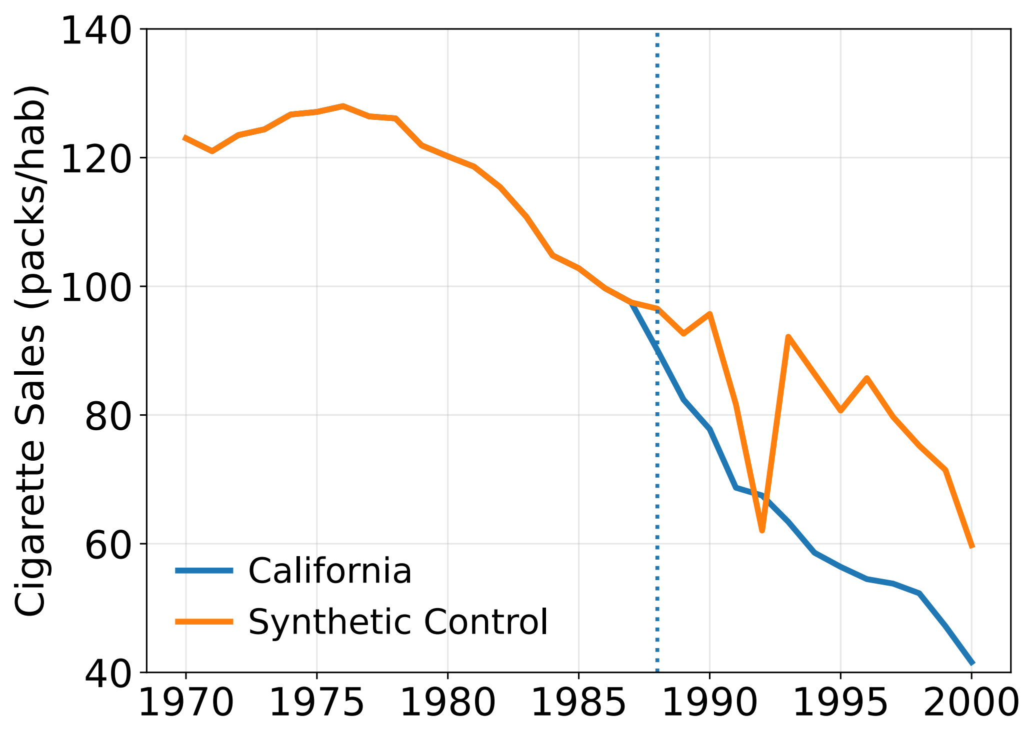
$$X_j = \begin{pmatrix} Y_{j,1} \\ .. \\ Y_{j,T_0} \\ Z_{j,1} \\ .. \\ Z_{j,T_0} \end{pmatrix}^T \in R^{2T_0}$$

Y cigarette sales, Z cigarette prices.

Model: $\underbrace{X_{\text{treat}}}_{p\times 1} \sim \underbrace{X_{\text{control}}}_{p\times n_0} \underbrace{w}_{n_0}$

Prediction: $\hat{Y}_{\text{synth}} = (Y_{t,j})_{\substack{t=1..T \\ j=2..n_0+1}} w$

😭 **Overfitting**



Synthetic controls: How to choose the predictor weights V?

1. Don't choose: set $V = I_p$, ie. $\|X\|_V = \|X\|_2$.
2. Rescale by the variance of the predictors:
   $V = \operatorname{diag}\left(\operatorname{var}\left(Y_{j,1}\right)^{-1}, .., \operatorname{var}\left(Y_{j,T_0}\right)^{-1}, \operatorname{var}\left(Z_{j,1}\right)^{-1}, .., \operatorname{var}\left(Z_{j,T_0}\right)^{-1}\right)$.
3. Minimize the pre-treatment mean squared prediction error (MSPE) of the treated unit:

Synthetic controls: How to choose the predictor weights V?

1. Don't choose: set $V = I_p$, ie. $\|X\|_V = \|X\|_2$.
2. Rescale by the variance of the predictors:
   $V = \text{diag}\left(\text{var}\left(Y_{j,1}\right)^{-1}, .., \text{var}\left(Y_{j,T_0}\right)^{-1}, \text{var}\left(Z_{j,1}\right)^{-1}, .., \text{var}\left(Z_{j,T_0}\right)^{-1}\right)$.
3. Minimize the pre-treatment mean squared prediction error (MSPE) of the treated unit:

Synthetic controls: How to choose the predictor weights V?

1. Don't choose: set $V = I_p$, ie. $\|X\|_V = \|X\|_2$.
2. Rescale by the variance of the predictors:
   $V = \text{diag}\left(\text{var}\left(Y_{j,1}\right)^{-1}, .., \text{var}\left(Y_{j,T_0}\right)^{-1}, \text{var}\left(Z_{j,1}\right)^{-1}, .., \text{var}\left(Z_{j,T_0}\right)^{-1}\right)$.
3. Minimize the pre-treatment mean squared prediction error (MSPE) of the treated unit:

$$\text{MSPE}(V) = \sum_{t=1}^{T_0} \left[Y_{1,t} - \sum_{j=2}^{n_0+1} w_j^*(V) Y_{j,t}\right]^2$$

$$= \| \left(Y_{1,t}\right)_{t=1..T_0} - \left(Y_{j,t}\right)^T_{\substack{j=2..n_0+1 \\ t=1..T_0}} \hat{w} \|_2^2$$

This solution is solved by running two optimization problems:

- Inner loop solving $w^*(V) = \text{argmin}_w \|X_{\text{treat}} - X_{\text{control}} w\|_V^2$
- Outer loop solving $V^* = \text{argmin}_V \text{ MSPE}(V)$

Synthetic controls: estimation without the outer optimization problem

Same coviarates: $X_j = \begin{pmatrix} Y_{j,1} \\ .. \\ Y_{j,T_0} \\ Z_{j,1} \\ .. \\ Z_{j,T_0} \end{pmatrix}^T$

Y cigarette sales, Z cigarette prices.

SCM minization with $V = I_p$, hence,
$\|X\|_V = \|X\|_2$.

$$w^* = \text{argmin}_w \ \|X_{\text{treat}} - X_{\text{control}} w\|_2^2$$

$$s.t. \ w_j \geq 0,$$

$$\sum_{j=2}^{n_0+1} w_j = 1$$

Synthetic controls: estimation without the outer optimization problem

Same coviarates: $X_j = \begin{pmatrix} Y_{j,1} \\ .. \\ Y_{j,T_0} \\ Z_{j,1} \\ .. \\ Z_{j,T_0} \end{pmatrix}^T$

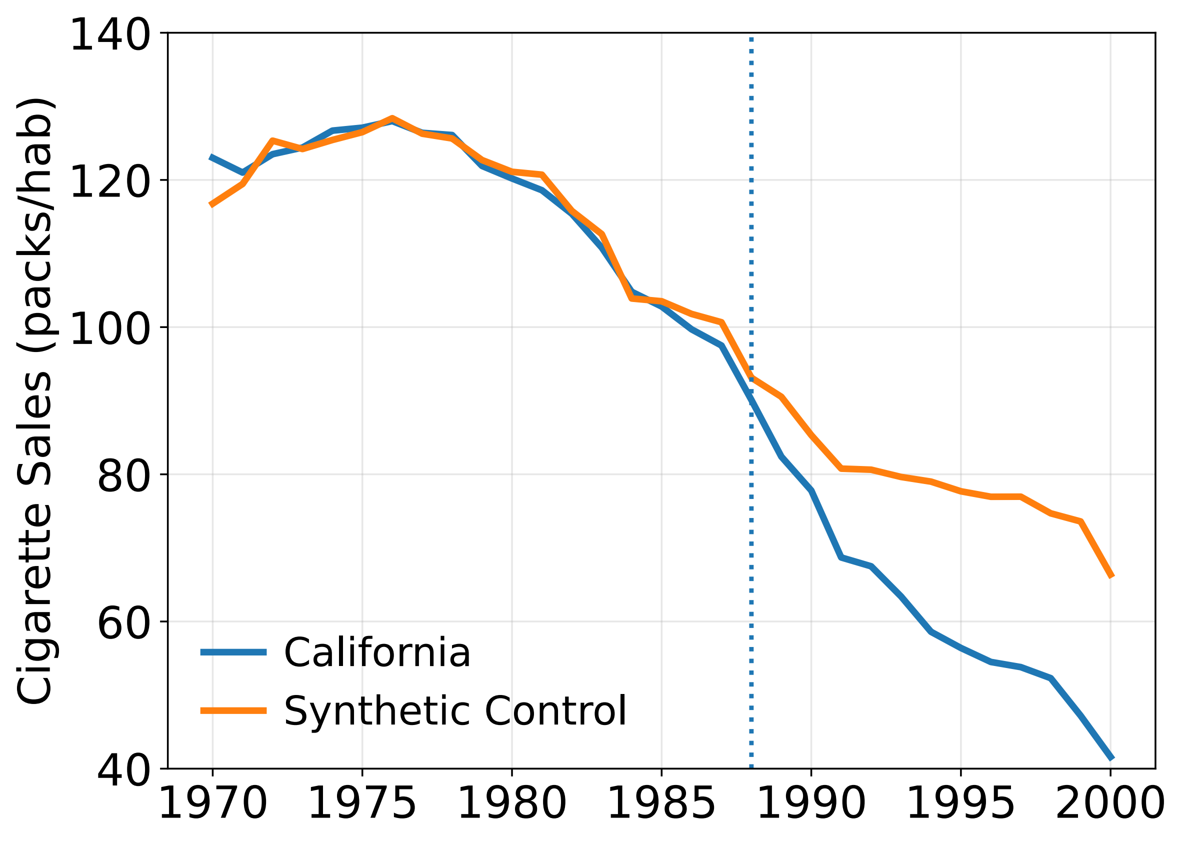
Y cigarette sales, Z cigarette prices.

SCM minization with $V = I_p$, hence, $\|X\|_V = \|X\|_2$.

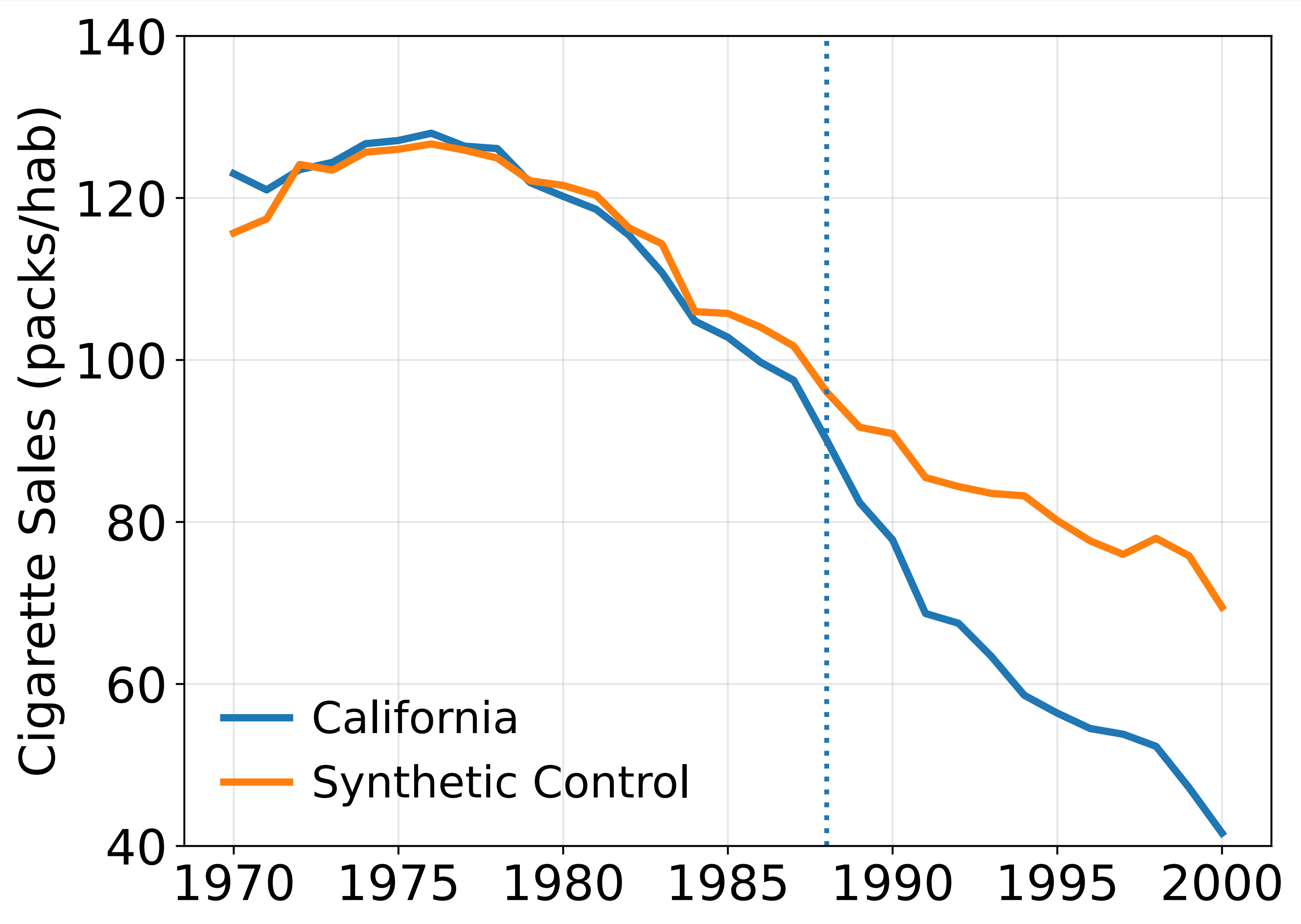
$w^* = \text{argmin}_w \; \|X_{\text{treat}} - X_{\text{control}} w\|_2^2$

$s.t. \; w_j \geq 0,$

$\sum_{j=2}^{n_0+1} w_j = 1$



Synthetic controls: estimation with the outer optimization problem



Synthetic controls: inference

Variability does not come from the variability of the outcomes

Indeed, aggregates are often not very noisy (once deseasonalized)...

Synthetic controls: inference

Variability does not come from the variability of the outcomes

Indeed, aggregates are often not very noisy (once deseasonalized)…

… but from the variability of the chosen control units

Treatment assignment introduces more noise than outcome variability.

Synthetic controls: inference

Variability does not come from the variability of the outcomes

Indeed, aggregates are often not very noisy (once deseasonalized)...

... but from the variability of the chosen control units

Treatment assignment introduces more noise than outcome variability.

(Abadie et al., 2010) introduced the placebo test to assess the variability of the synthetic control.

There is also a modern approach on inference for SCM based on Conformal prediction (Chernozhukov et al., 2021) (see end of the slides for intuition).

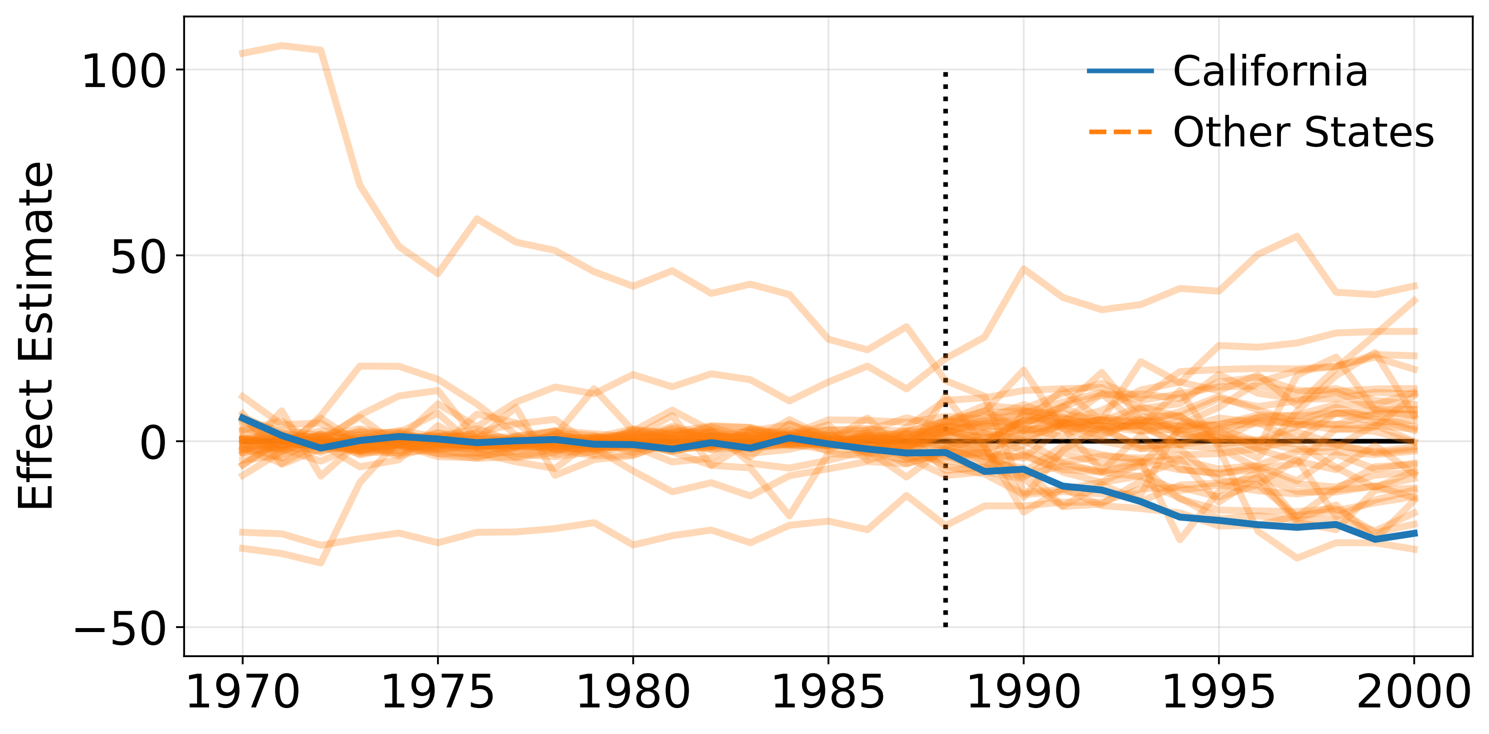
Synthetic controls: inference with Placebo tests

Idea of placebo tests, also called Fisher's Exact tests

- Permute the treated and control exhaustively.
- For each unit, we pretend it is the treated while the others are the control: we call it a placebo
- Compute the synthetic control for each placebo: it should be close to zero.

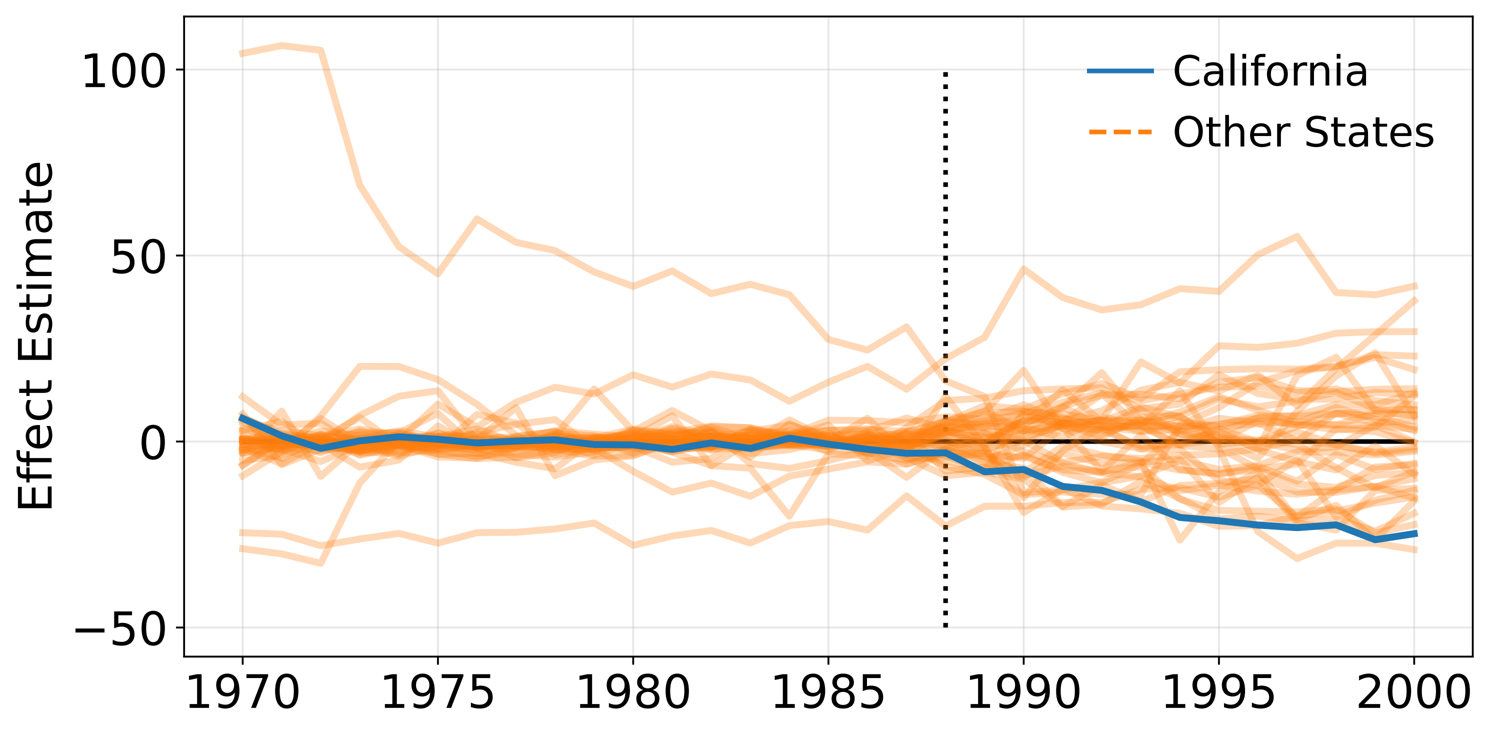
Synthetic controls: inference with Placebo tests, example

Placebo estimation for all 38 control states



Synthetic controls: inference with Placebo tests, example

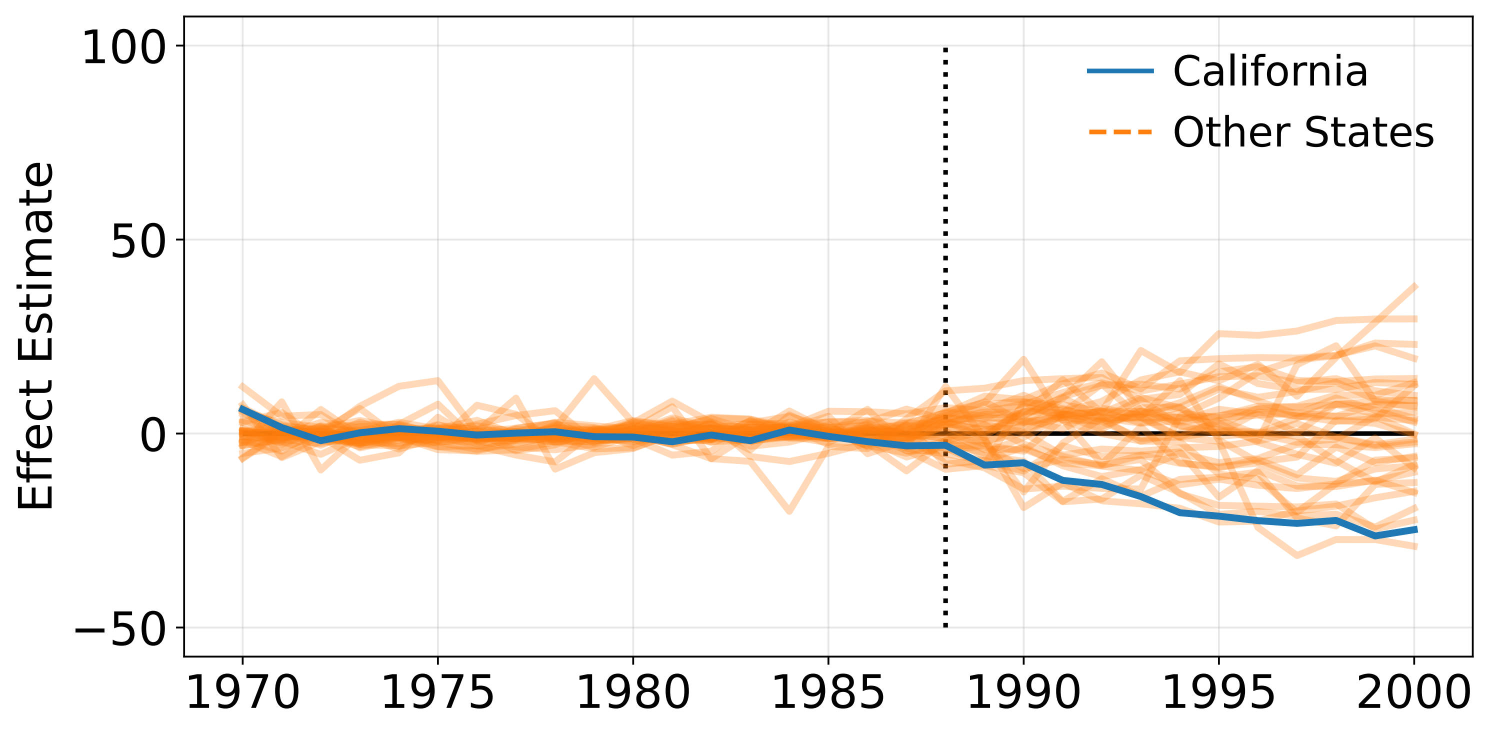
Placebo estimation for all 38 control states



- More variance after the treatment for California than before.
- Some states have pre-treatment trends which are hard to predict.

Synthetic controls: inference with Placebo tests, example

Placebo estimation for 34 control states with "good" pre-treatment fit

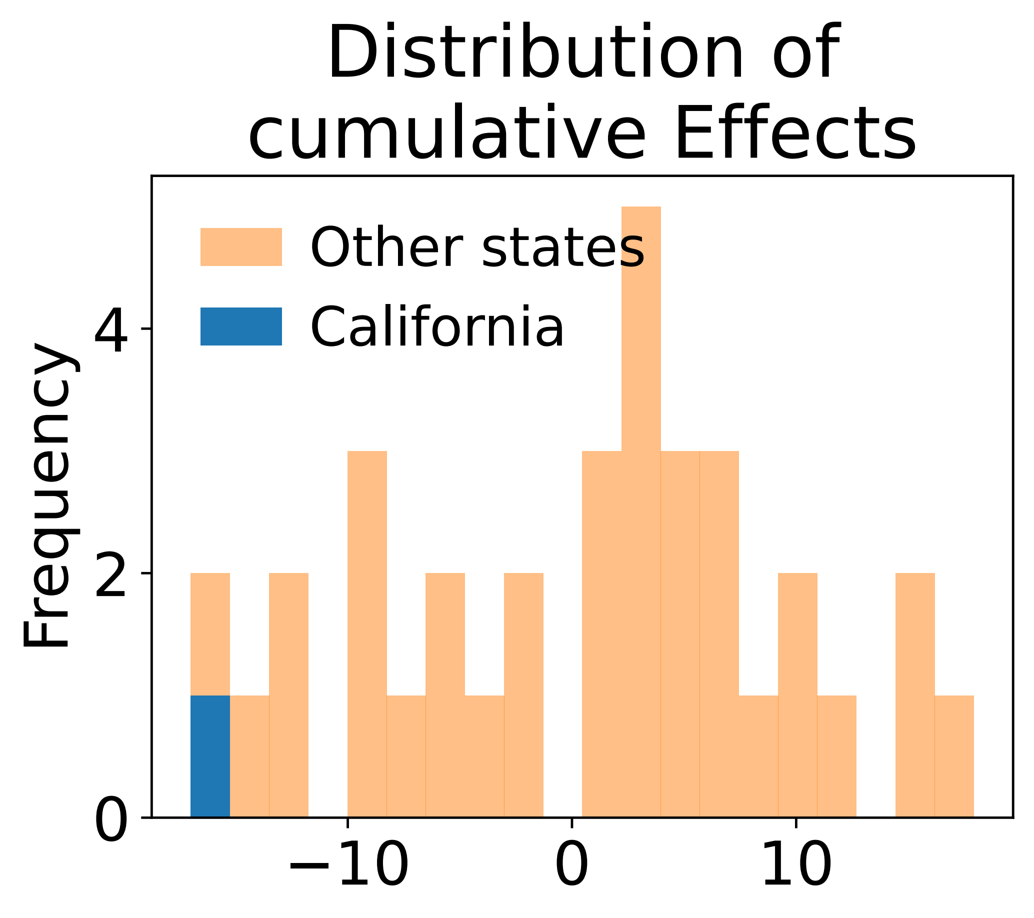


I removed the states above the 90 percentiles of the distribution of the pre-treatment fit.

Synthetic controls: inference with Placebo tests, example

California absolute cumulative effect

$\hat{\tau}_{\text{scm, california}} = -17.00$



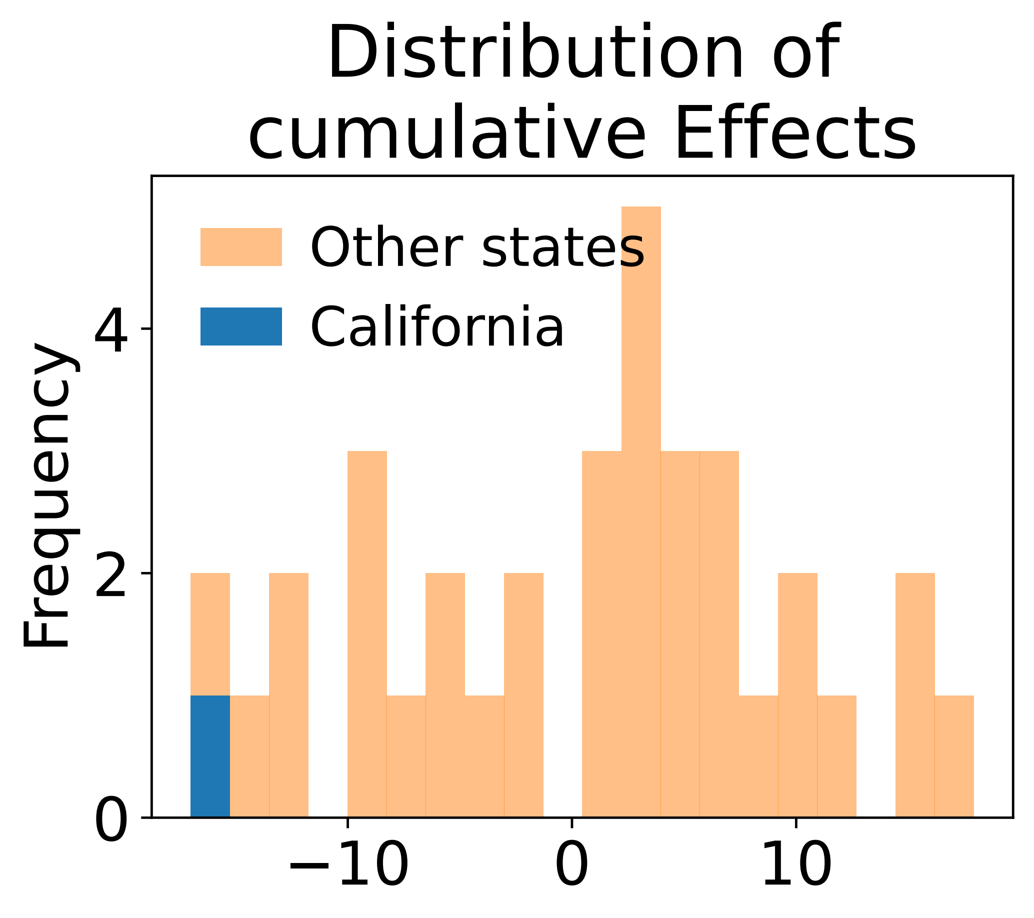
Synthetic controls: inference with Placebo tests, example

California absolute cumulative effect

$$\hat{\tau}_{\text{scm, california}} = -17.00$$

Get a p-value

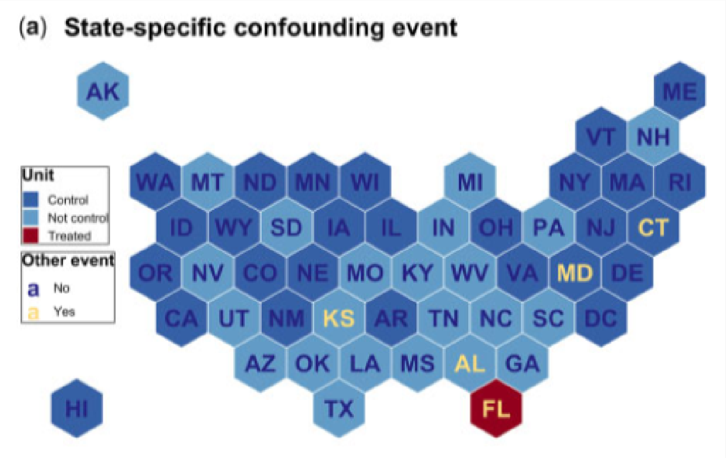
$$\text{PV} = \frac{1}{n_0} \sum_{j=2}^{n_0} \mathbb{1}\left(\left|\hat{\tau}_{\text{scm, california}}\right| > \left|\hat{\tau}_{\text{scm},j}\right|\right)$$
$$= 0.029$$



Synthetic controls failure: confounding event for some controls

Common causes of outcome and for only part of the controls and the treated unit

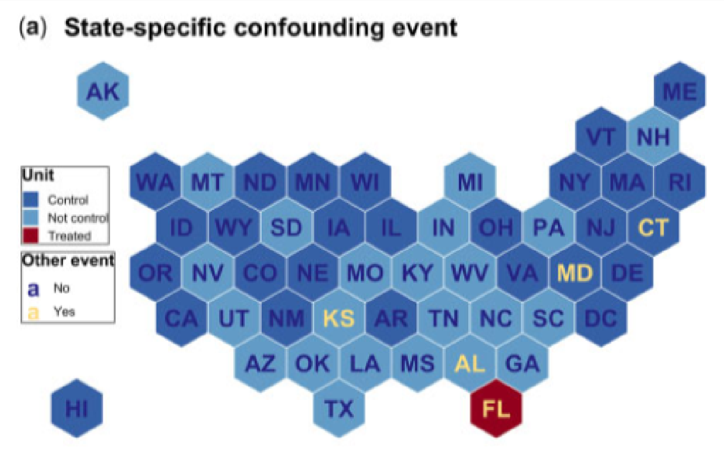
TODO: explain setup (Degli Esposti et al., 2020)



Synthetic controls failure: confounding event for some controls

Common causes of outcome and for only part of the controls and the treated unit

TODO: explain setup (Degli Esposti et al., 2020)

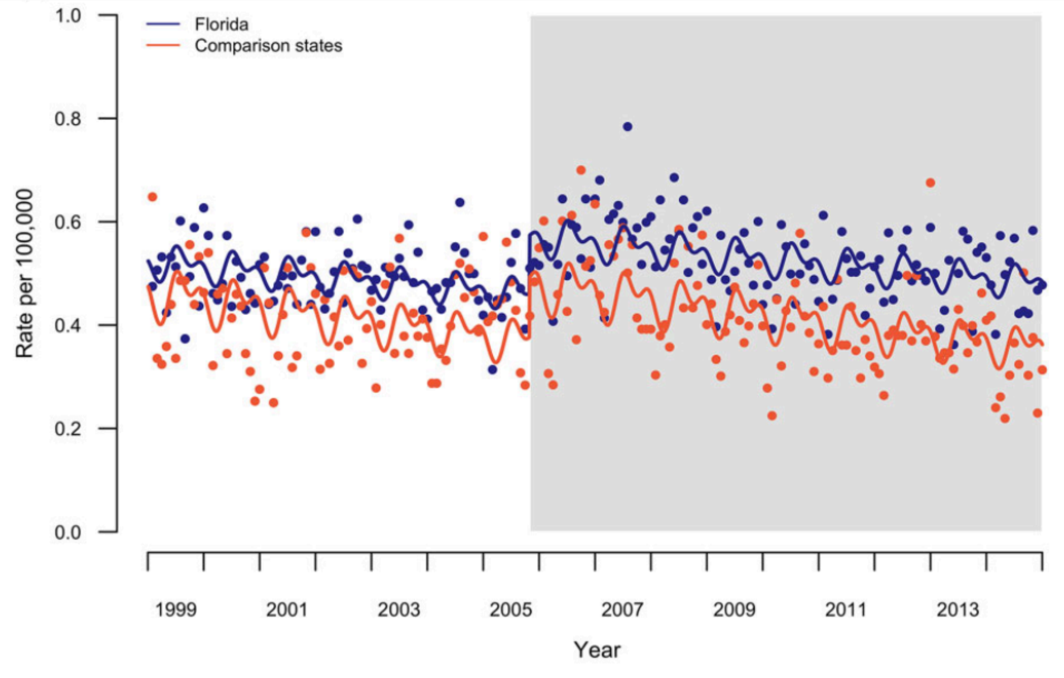


Suppose that this other event have an impact on the outcome after the treatment.

For state in [KS, MD, AL, CT, FL], there is a step change in the outcome after the treatment: $\mathbb{1}[t > T_0]$

Synthetic controls failure: appropriate controls

Focus only on states affected by the confounding events

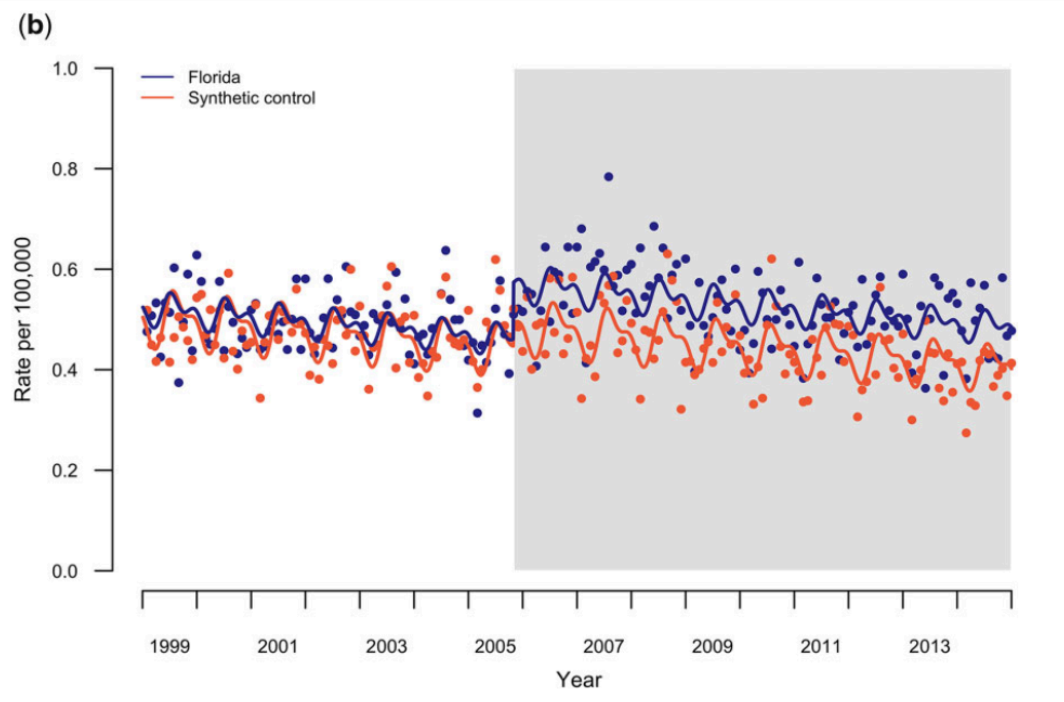


Here, the comparison states are: KS, MD, AL, CT : also affected by the counfounding event.

No problem: we would conclude to no effect of the treatment.

Synthetic controls failure: data-driven controls

Focus on all comparison states



SCM matches pre-treatment trends, without taking into account the confounding event.

Problem: we would falsely conclude to a positive treatment effect.

Synthetic controls: Take-away

Pros

- More convincing for parallel trends assumption.
- Handle multiple time periods.
- Data driven.
- Gives confidence intervals thanks to placebo test.

Synthetic controls: Take-away

Pros

- More convincing for parallel trends assumption.
- Handle multiple time periods.
- Data driven.
- Gives confidence intervals thanks to placebo test.

Cons

- Requires many control units to yield good pre-treatment fits.
- Might be prone to overfitting during the pre-treatment period.
- Still requires a strong assumption: the weights should also balance the post-treatment unexposed outcomes ie. conditional ignorability. See (Arkhangelsky et al., 2021) for discussions.
- Still requires the no-anticipation assumption.

Synthetic controls: Take-away

Pros

- More convincing for parallel trends assumption.
- Handle multiple time periods.
- Data driven.
- Gives confidence intervals thanks to placebo test.

Cons

- Requires many control units to yield good pre-treatment fits.
- Might be prone to overfitting during the pre-treatment period.
- Still requires a strong assumption: the weights should also balance the post-treatment unexposed outcomes ie. conditional ignorability. See (Arkhangelsky et al., 2021) for discussions.
- Still requires the no-anticipation assumption.

Synthetic controls: Take-away

Pros

- More convincing for parallel trends assumption.
- Handle multiple time periods.
- Data driven.
- Gives confidence intervals thanks to placebo test.

Cons

- Requires many control units to yield good pre-treatment fits.
- Might be prone to overfitting during the pre-treatment period.
- Still requires a strong assumption: the weights should also balance the post-treatment unexposed outcomes ie. conditional ignorability. See (Arkhangelsky et al., 2021) for discussions.
- Still requires the no-anticipation assumption.

Interrupted time-series: methods without a control group

Interrupted Time Series: intuition

Setup

- One treated unit, no control unit.
- Multiple time periods.
- Sometimes, predictors are availables: there are called exogeneous covariates.

Interrupted Time Series: intuition

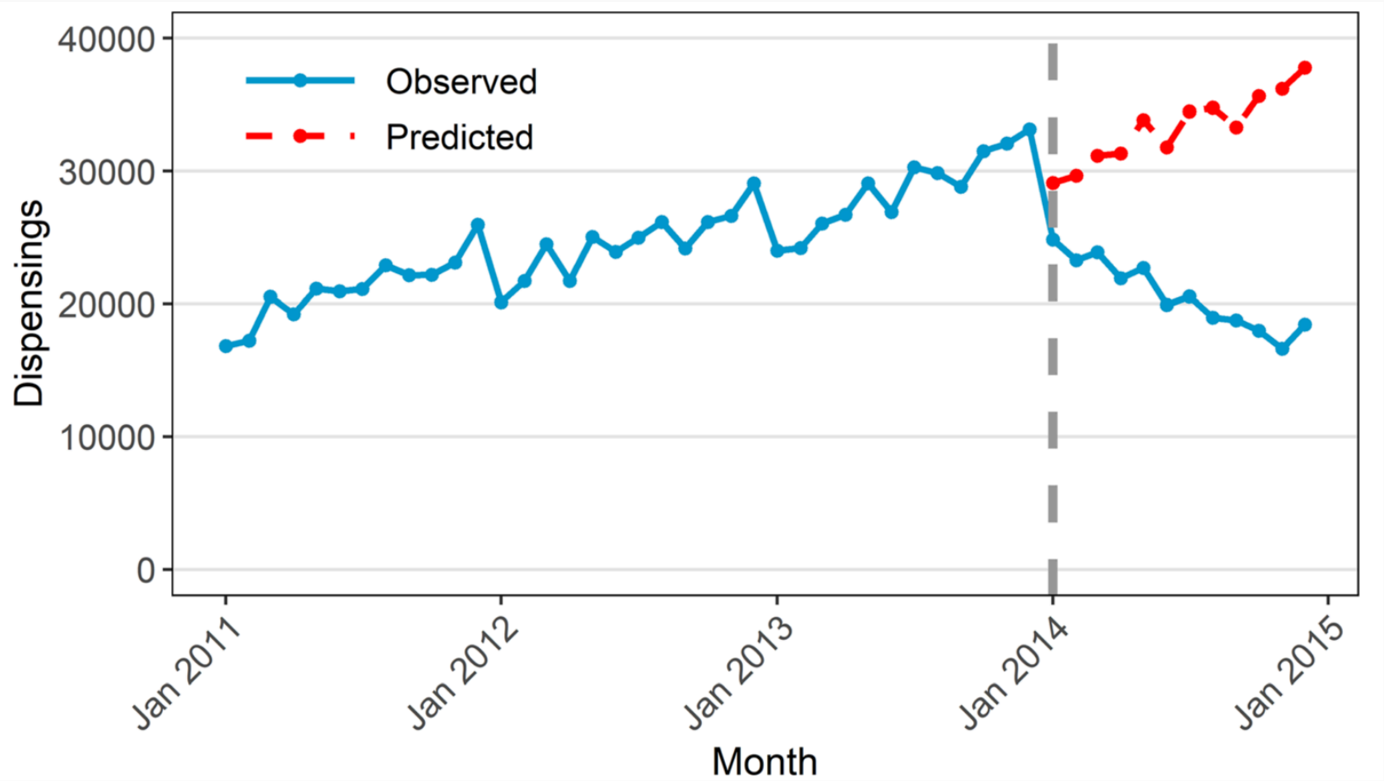
Setup

- One treated unit, no control unit.
- Multiple time periods.
- Sometimes, predictors are availables: there are called exogeneous covariates.

Intuition

- Model the pre-treatment trend: $Y_{t(1)}$ for $t < T_0$
- Predict post-treatment trend as the control: $\hat{Y}_t(0)$ for $t > T_0$
- Obtain treatment effect by taking the difference between observed and predicted post-treatment observations: $Y_t(1) - \hat{Y}_t(0)$

Interrupted Time Series: illustration from (Schaffer et al., 2021)



Y_t: Dispensations of quetiapine, an anti-psychotic medicine.

Treatment: Restriction of the conditions under which quetiapine could be subsidised.

Tools

- ARIMA models: AutoRegressive Integrated Moving Average

Motivation of ARIMA

- Structure of autodependance between observation (auto-regression, moving average),
- Linear trends,
- Seasonality.

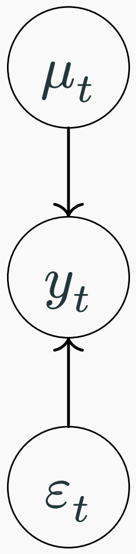
Good reference

Forecasting (fpp3): Principles and Practice, chapter 8

ARIMA are State Space Models (SSM) says the machine learning community

What is a state space model?

- The time series has two components: the state μ_t and the observation y_t.
- The state is a latent variable that evolves over time.
- The observation is a noisy version of the state: $y_t = \mu_t + \varepsilon_t$



ARIMA are State Space Models (SSM) says the machine learning community

ARIMA are State Space Models (SSM) says the machine learning community

What is a state space model?

- The time series has two components: the state μ_t and the observation y_t.
- The state is a latent variable that evolves over time.
- The observation is a noisy version of the state: $y_t = \mu_t + \varepsilon_t$

Why showing this formulation ?

- I better understand ARIMA formulated as state space models.
- SSM are more general than ARIMA models.
- ARIMA are (often) fitted with SSM optimization algorithms.

Good reference

(Murphy, 2022, book 2, chap 29)

State space models: AR(1) or AR(2) model example

AR(1)

DAG

$$y_{t-1} \to y_t \to y_{t+1}$$

Formalization

Observation: $y_t = \rho y_{t-1} + \varepsilon_{y,t}$

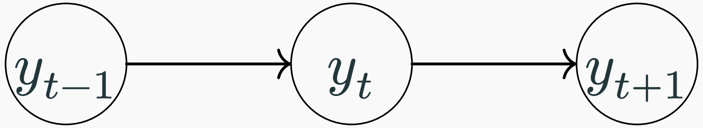
with $\varepsilon_{y,t} \sim N(0, \sigma_y^2)$

$$|\rho| < 1$$

Auto-regression time series model an outcome as a linear regression of its prior values.

AR(1)

DAG



Formalization

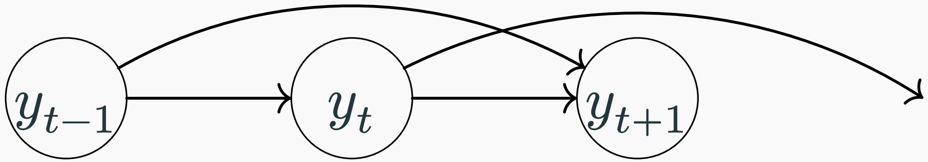
Observation: $y_t = \rho y_{t-1} + \varepsilon_{y,t}$

with $\varepsilon_{y,t} \sim N\left(0, \sigma_y^2\right)$

$$|\rho| < 1$$

AR(2)

DAG



Formalization

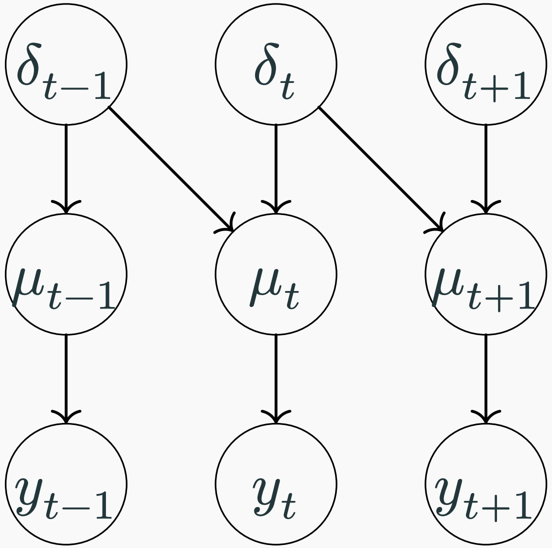
Observation: $y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_{y,t}$

with $\varepsilon_{y,t} \sim N\left(0, \sigma_y^2\right)$

$$|\rho_1| < 1, |\rho_2| < 1$$

State space models: MA(1) ie. ARIMA(0,0,1) model example

DAG



Formalization

Observation: $y_t = \mu_t + \theta\mu_{t-1} + \varepsilon_{y,t}$

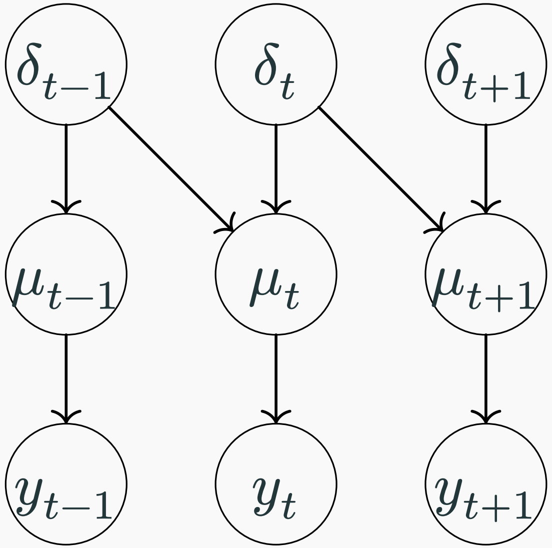
Latent: $\mu_t = \delta_t$

with $\varepsilon_{y,t} \sim N(0, \sigma_y^2)$

$$\delta_t \sim N(0, \sigma_\delta^2)$$

State space models: MA(1) ie. ARIMA(0,0,1) model example

DAG



Formalization

Observation: $y_t = \mu_t + \theta\mu_{t-1} + \varepsilon_{y,t}$

Latent: $\mu_t = \delta_t$

with $\varepsilon_{y,t} \sim N(0, \sigma_y^2)$

$$\delta_t \sim N(0, \sigma_\delta^2)$$

The MA time series models the residual of the regression of y_t on its previous values as a linear combination of the previous residuals : ie. vanishing shocks.

State space models: ARMA(p, q) ie. ARIMA(p,0,q) model example

TODO: check the SSM formulation

Formalization

Observation: $y_t = \mu_t + \varepsilon_{y,t}$

Latent: $\mu_t = \delta_t + \theta\delta_{t-1} + \varepsilon_{\mu,t}$

with $\varepsilon_{y,t} \sim N\left(0, \sigma_y^2\right)$

$$\varepsilon_{\mu,t} \sim N\left(0, \sigma_\mu^2\right)$$

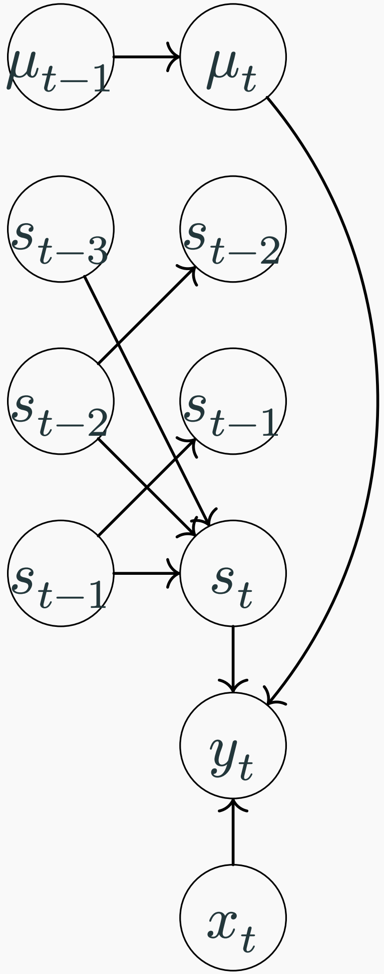
$$\delta_{\mu,t} \sim N\left(0, \sigma_\delta^2\right)$$

Unfolding the state space equations

$y_t = \sum_{i=1}^{p} \rho_i y_{t-i} + \sum_{j=1}^{q} \theta_j \delta_{t-j} + \varepsilon_{y,t}$

State space models: Adding a seasonnality and a covariate component

DAG



Formalization

Observation with covariates and seasonality:

$$y_t = \mu_t + \beta x_t + s_t + \varepsilon_{y,t}$$

Where seasonality:

$$s_t = -\sum_{k=1}^{S-1} s_{t-k} + \varepsilon_{s,t}$$

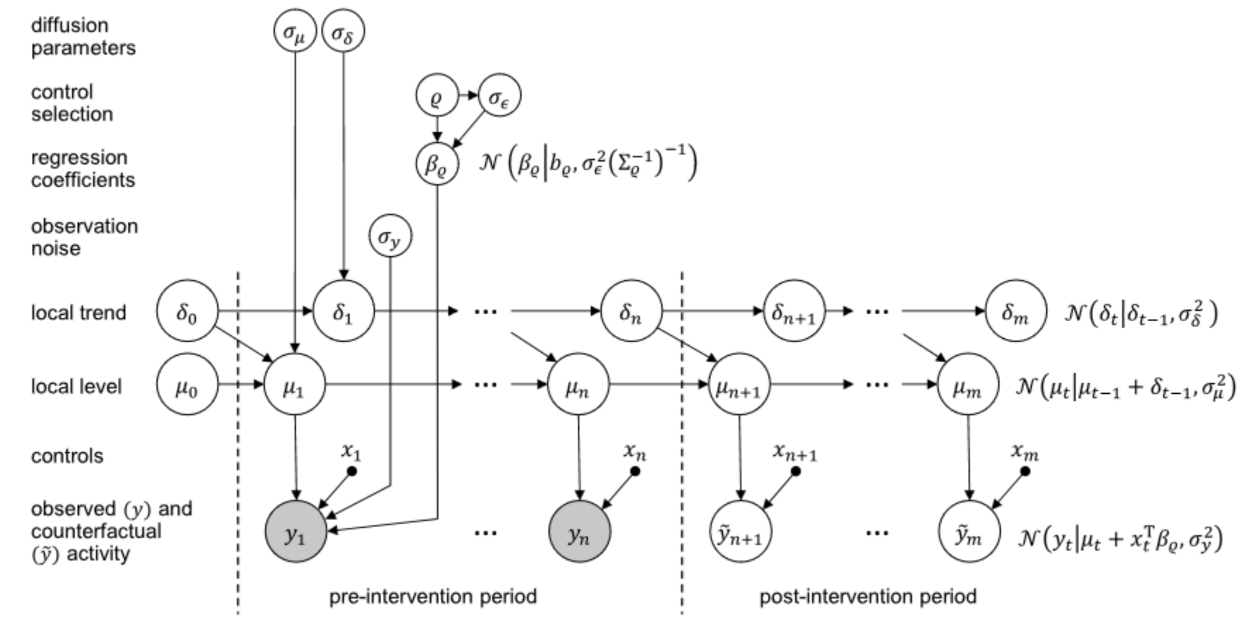
with $\varepsilon_s, t \sim N(0, \sigma_s^2)$

State space models: General formulation

SSM have a more general formulation than ARIMA models

- State equation: $\alpha_t = T_t\alpha_{t-1} + c_t R_t \eta_t$ with $\eta_t \sim N(0, Q_t)$
- Observation equation: $y_t = Z_t\alpha_t + \beta^T x_t + H_t\varepsilon_t$ with $\varepsilon_t \sim N(0, V_t)$
- η_t and ε_t are white noise terms.

Complex SSM DAG from the Causal Impact paper (Brodersen et al., 2015)



State space models: a brief word on fitting (ie. learning the parameters)

When the error terms are gaussians

These modeles are called linear Gaussian state space model (LG-SSM) or linear dynamical system (LDS).

The likelihood is jointly gaussian

Closed form formula for the likelihood of the data under the model.

State space models: a brief word on fitting (ie. learning the parameters)

When the error terms are gaussians

These modeles are called linear Gaussian state space model (LG-SSM) or linear dynamical system (LDS).

The likelihood is jointly gaussian

Closed form formula for the likelihood of the data under the model.

Expectation-Minimization: a widespread algorithm for fitting

- Expectaction: Compute the joint likelihood of the data and the parameters (observed outcome, unknown state) given the parameters.
- Maximization: find parameters maximizing the likelihood: analytically since gaussian.
- Iter until convergence to a (local) maximum of likelihood.

Modern state space models

- Long Short Term Memory (LSTM) networks (Graves & Graves, 2012): a type of Recurrent Neural Network (RNN) that can learn long-term dependencies. Was state of the art for language tasks before transformers.
- Mamba (Gu & Dao, 2023): A recent proposition to mitigate the main limitations of transfomers which is high complexity relative to the length of the sequence. Good blog-style introduction in (Ayonrinde, 2024).

Example of ITS with ARIMA: the French antibiotics campaign of 2002-2007

Context

In 2001, compared to the European Union countries, France was a country where:

- the population consumed the most antibiotics in town
- the resistance of Streptococcus pneumoniae to β-lactams was the highest (53%)
- a significant number of antibiotic prescriptions would be unnecessary (viral infections)

Campaign (october 2002)

France implemented a national plan to "preserve the effectiveness of antibiotics and improve their use" with the main action undertaken by the National Health Insurance.

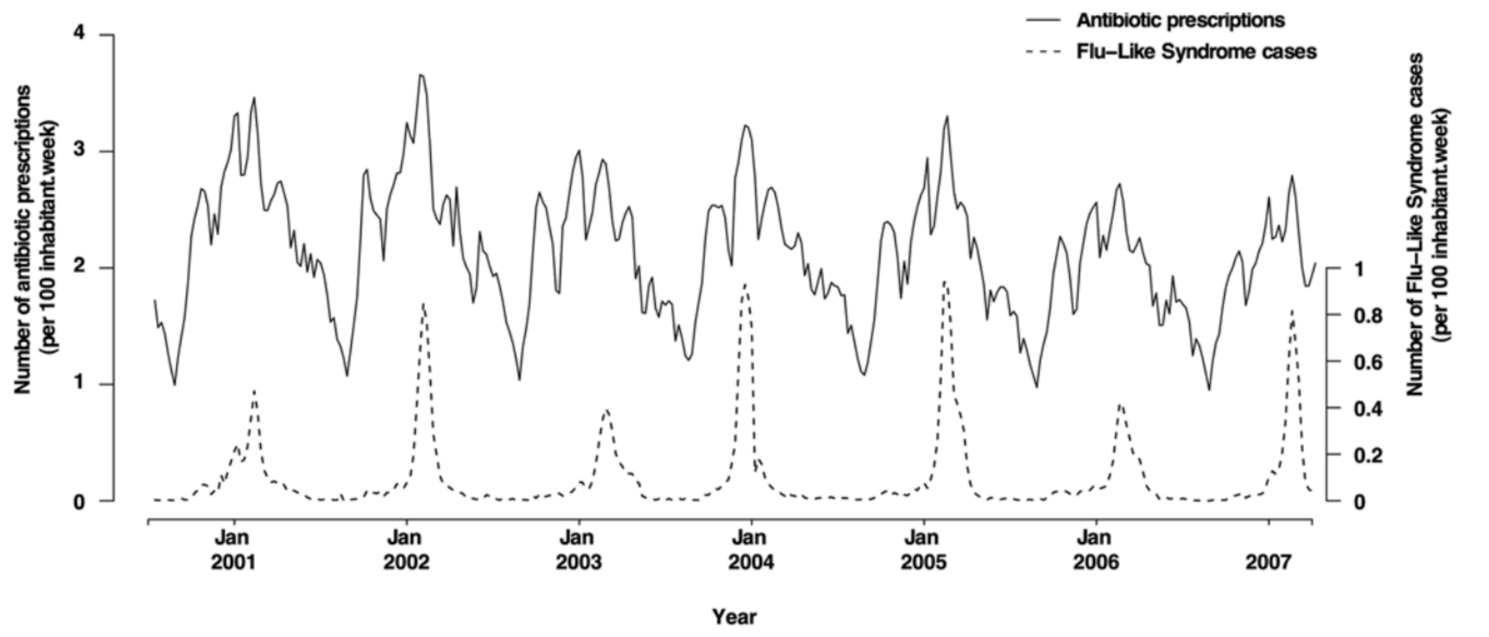
The campaign was reactivated every year until from october to march.

Question

What has been the effect of the campaign on the consumption of antibiotics? (Sabuncu et al., 2009)

Example of ITS with ARIMA: the French antibiotics campaign of 2002-2007

Weekly reimbursed prescription of antibiotics in town



Interventions during the months of october to march: $\mathrm{month}(t) \in M_0$.

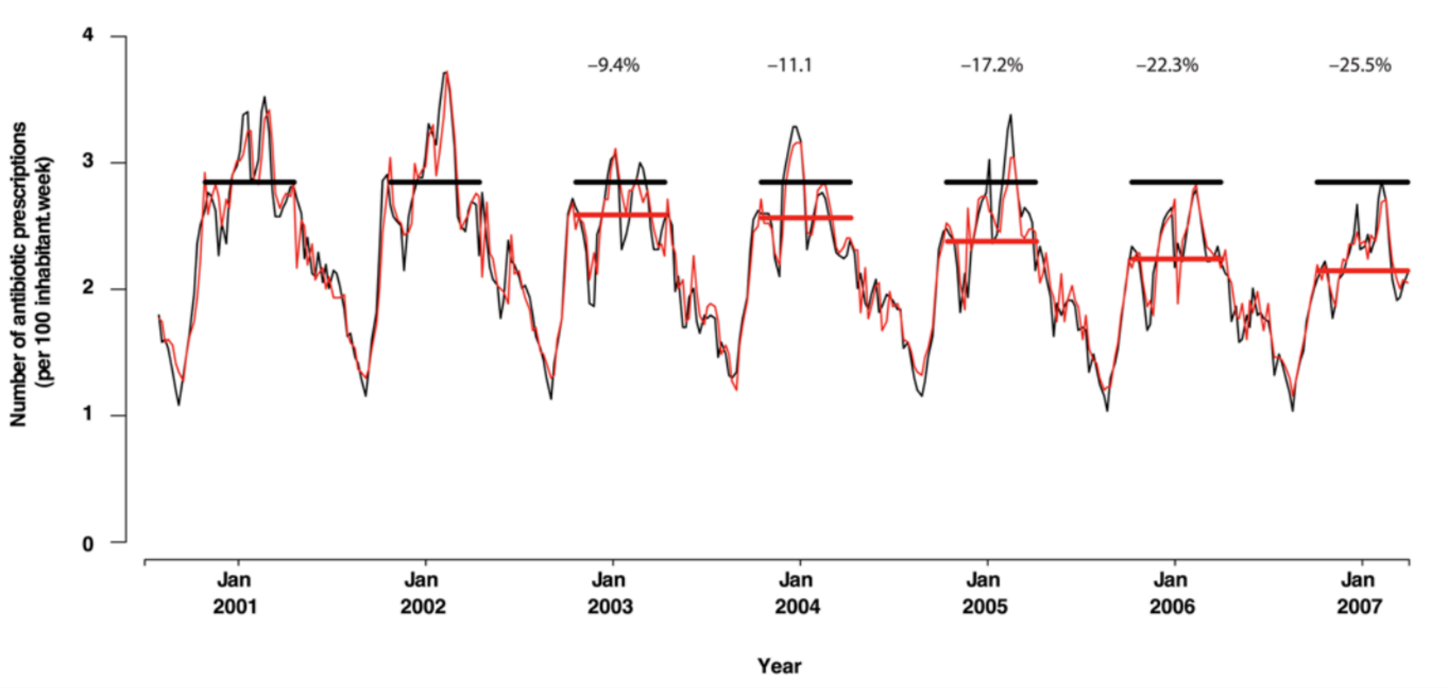
Estimation

- Fit an ARIMA model on the pre-treatment trend
- Introduce an additive term for the intervention:

$$Y_t = c + \sum_i \hat{\tau}_i \mathbb{1}[\text{month}(t) \in M_0 \land \text{year}(t) == i] + \underbrace{[a(B)^{-1} - b(B)\varepsilon_s]}_{\text{ARIMA term fitted on pre-treatment}}$$

- Assess if the additive term and other parameters are significantly different pre-treatment and post-treatment.

Example of ITS with ARIMA: the French antibiotics campaign of 2002-2007



- Red curve: arima fitted with intervention
- Red Horizontal line: intervention effect fitted during intervention
- Black curve: arima fitted without intervention
- Black horizontal line: intervention effect fitted pre-intervention

Example of ITS with more general SSM: Causal impact

TODO

A word on model families for ITS

We saw ARIMA models and the more general class of state space models.

However, we could any model that we want to fit the pre-treatment trend !

A word on model families for ITS

We saw ARIMA models and the more general class of state space models.

However, we could any model that we want to fit the pre-treatment trend !

- Facebook prophet model (Taylor & Letham, 2018) uses Generalized Additive Models (GAM).
- Any sklearn estimator could do the trick: Linear regression, Random Forest, Gradient Boosting…

A word on model families for ITS

We saw ARIMA models and the more general class of state space models.

However, we could any model that we want to fit the pre-treatment trend !

- Facebook prophet model (Taylor & Letham, 2018) uses Generalized Additive Models (GAM).
- Any sklearn estimator could do the trick: Linear regression, Random Forest, Gradient Boosting…

A word on model families for ITS

We saw ARIMA models and the more general class of state space models.

However, we could any model that we want to fit the pre-treatment trend !

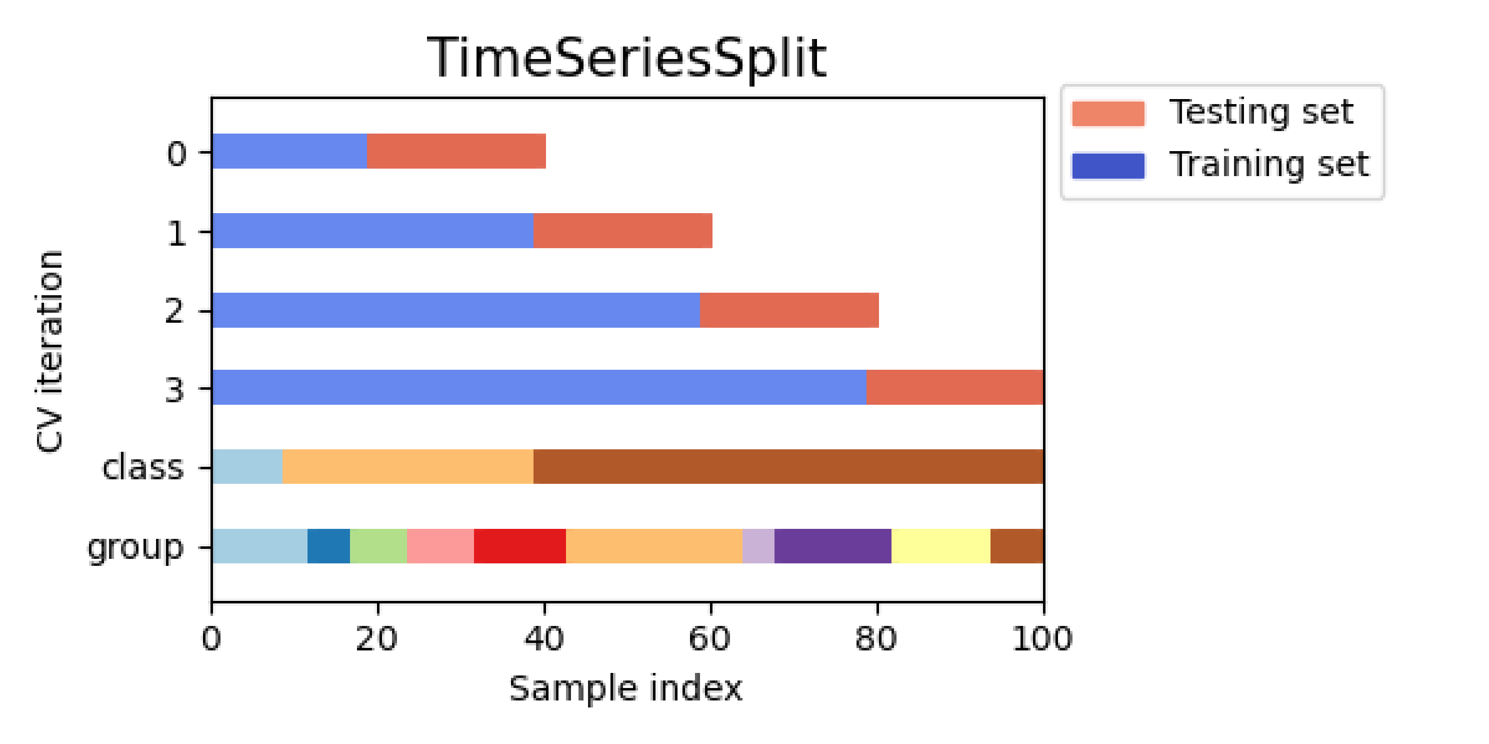
- Facebook prophet model (Taylor & Letham, 2018) uses Generalized Additive Models (GAM).
- Any sklearn estimator could do the trick: Linear regression, Random Forest, Gradient Boosting…

⚠️ You should pay attention to appropriate train/test split when cross-validating a time-series model not to use the future to predict the past.

Relevant remark for all time series models (even ARIMA or state space models).

Cross-validation for time-series models

```python
from sklearn.model_selection import TimeSeriesSplit
```



This avoids to use the future to predict the past.

Main threat to validity for an ITS: historical bias

⚠️ If there is a co-intervention, it will impact the outcome trend and bias the treatment effect estimation.

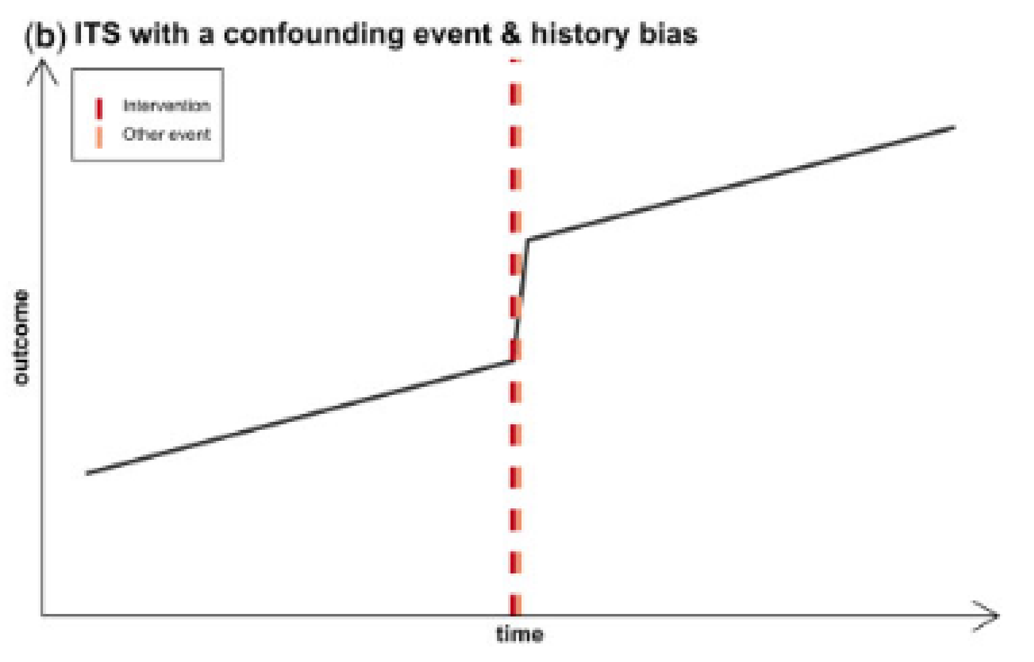


Illustration from (Degli Esposti et al., 2020, Fig. 1)

Main threat to validity for an ITS: historical bias

⚠️ If there is a co-intervention, it will impact the outcome trend and bias the treatment effect estimation.

💡 Adding a control series of predictors can help to mitigate this bias.

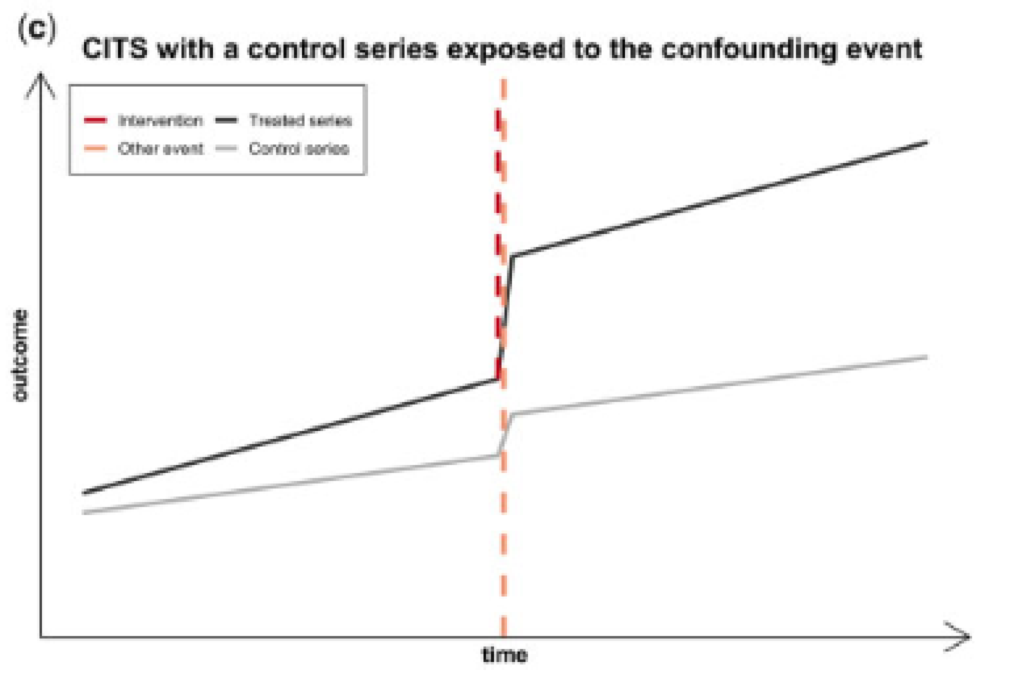


Illustration from (Degli Esposti et al., 2020, Fig. 1)

Take-away on ITS

Pros

- Suitable when no control unit is available. The pre-treatment trend is the control.
- Handles multiple time periods.
- A lot of software available (eg. ARIMA models).
- Simple: few parameters to tune.

Cons

- Prone to bias by other events happening around the treatment time and impacting the outcome trend.
- Prone to overfitting of the pre-treatment trend.

An attempt to map event study methods

Methods	Characteristics	Hypotheses	Community	Introduction	Good reference
DID/TWFE	Treated/control units, few time periods, no predictors	Parallel trends, no anticipation, prone to overfitting	Economics	Causal Inference for the Brave and True, chapter 13	(Arkhangelsky & Imbens, 2024)
ARIMA, ITS	No controls, no/few predictors, seasonality	Stationnarity , no anticipation, prone to overfitting	Epidemiology, Economics	Forecasting: Principles and Practice	(Schaffer et al., 2021)
State space models	Multiple time periods, control units or predictors, generalization of ARIMA	Contional ignorability on predictors, goodness of fit pre-treatment	Machine learning, bayesian methods	(Brockwell & Davis, 2016, chapter 9)	(Murphy, 2022, chapter 18)
Synthetic control	Treated/control units, multiple time periods	Conditional parallel trend on controls, goodness of fit pre-treatment	Economics	Causal Inference for the Brave and True	(Abadie, 2021)

A summary on R packages for event studies

Package name	Methods	Predictors	Control units	Multiple time periods
did	Difference-in-differences	❌	❌	❌
forecast	ARIMA, ITS	✅	❌	✅
Synth	Synthetic control	❌	✅	✅
Causal impact	Bayesian state space models	✅	❌	✅

A summary on Python packages for event studies

Package name	Methods	Predictors	Control units	Multiple time periods
statsmodels.OLS	Difference-in-differences, TWFE	❌	❌	❌
statsmodels	ARIMA(X), ITS, bayesian state space models	✅	❌	✅
pmdarima	ARIMA(X), ITS	✅	❌	✅
SyntheticControlMethods	Synthetic control	❌	✅	✅
pysyncon	Synthetic control	❌	✅	✅
causalimpact (pymc implementation)	Bayesian state space models	✅	❌	✅
causal-impact (statsmodels implementation)	Bayesian state space models	✅	❌	✅

Final word -- What methods to chose: some guides

DID-family methods

- Control units available (at least one)
- Few time periods
- Parallel trend is credible (if necessary by adjusting the model on predictors).

Synthetic Control Methods

- Mutiple and different controls as well as multiple time periods
- Pre-treatment outcomes of the control units predict well the treated unit outcome.
- No-spill over from the treatment to the control units.

ITS: SARIMA or state space models

- No evident control units
- Pre-treatment outcome of the treated unit seems a good control
- Control predictors not impacted by the treatment availables
- No co-intervention that could impact the treated outcome.

Python hands-on

To your notebooks 🧑‍💻 !

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

Supplementary materials

Synthetic controls: conformal prediction inference

Introduced by (Chernozhukov et al., 2021)

- Recast the problem as counterfactual inference, ie. predict: $Y_{it}(0)$ for $t > T_0$
- Test hypothesis: H_0 eg. $H_0 = (0, 0, .., 0)$ ie no effect for $t > T_0$
- This imply the generation of a hypothesis counterfactual trajectory $Y_t(0)$

Question

Are the post-treatment residuals of a model fitted on the hypothesis counterfactual trajectory an outlier of the distribution of the residuals pre-treatment?

Why does this works?

Syntehtic controls estimation are invariant under the time series dimension so we can resample under this dimension to introduce data variability.

Conformal inference: hypothesis generation

- Test a hypothesis : H_0 eg. $H_0 = (0, 0, .., 0)$ ie no effect for $t > T_0$
- Gerenate a counterfactual trajectory $Y_t(0)$ under this null

Conformal inference: Fit a model and compute residuals

- Fit a counterfactual model on the full generated trajectory: $\hat{Y}_t$
- Compute the residuals: $\hat{u}_t = Y_t(0) - \hat{Y}_t$

Conformal inference: test statistic and resampling

Summarize the residuals in a statistic: $S(\hat{u}) = \left(\frac{1}{\sqrt{T-T_0+1}} \sum_{t=T_0+1}^{T} |\hat{u}_t|^q\right)^{\frac{1}{q}}$

Conformal inference: resampling

Resample this statistic by block permutation π of the time periods

Same as permuting the data since SCM are invariant under the time series dimension.

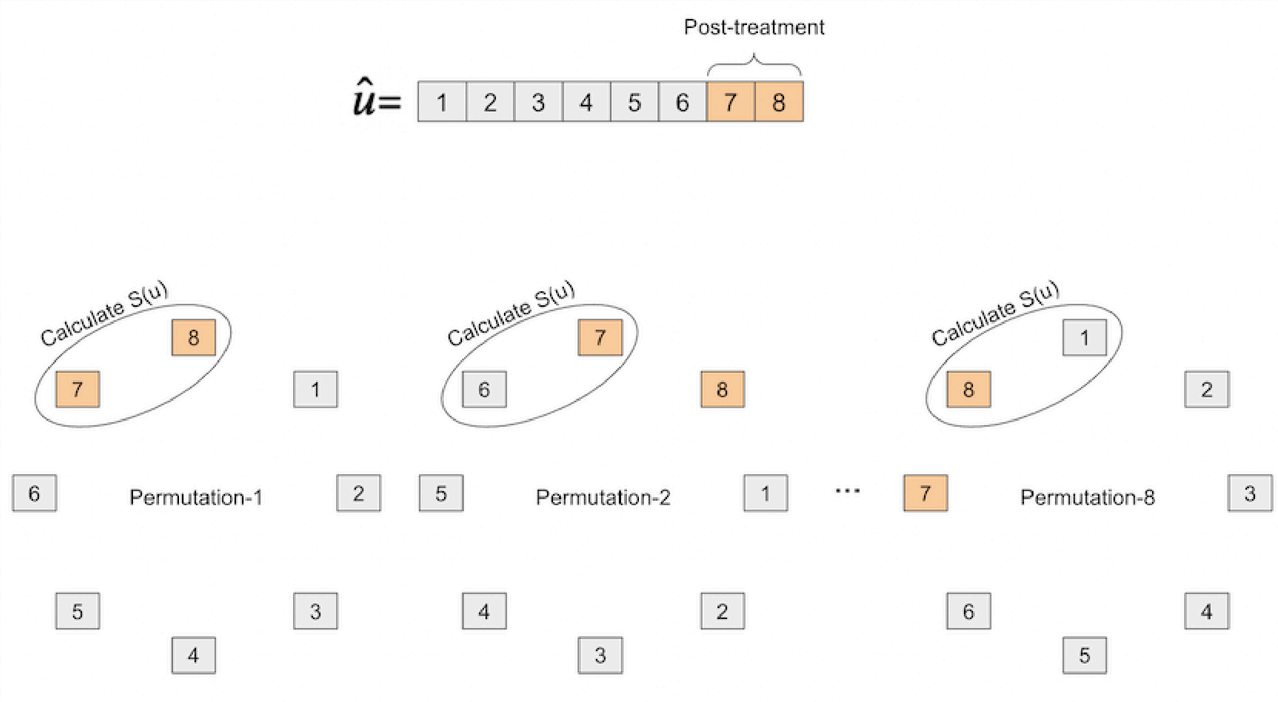


Image from: Causal Inference for the Brave and True

Conformal inference: P-value

- Assess if the post-treatment statistics is an outlier of this distribution.
- P-value: $\hat{F}(x) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{1}\left[S\left(\hat{u}_{\pi_0}\right) \leq S(\hat{u}_\pi)\right]$ where π_0 is the original data.

Conformal inference: confidence intervals

TODO

Conditional difference-in-differences

TODO

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.

Arkhangelsky, D., & Imbens, G. (2024). Causal models for longitudinal and panel data: A survey. The Econometrics Journal, 27(3), C1–C61.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. American Economic Review, 111(12), 4088–4118.

Bibliography

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

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. Journal of Economic Perspectives, 31(2), 3–32.

Ayonrinde, K. (2024). Mamba Explained. The Gradient.

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.

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.

Brehm, M. E., Brehm, P. A., & Saavedra, M. (2022). The Ohio vaccine lottery and starting vaccination rates. American Journal of Health Economics, 8(3), 387–411.

Bibliography

Brockwell, P. J., & Davis, R. A. (2016). Introduction to time series and forecasting, Third edition. Springer. http://repository.cinec.edu/bitstream/cinec20/1109/1/2016_Book_IntroductionToTimeSeriesAndFor.pdf

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models.

Chernozhukov, V., Wüthrich, K., & Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. Journal of the American Statistical Association, 116(536), 1849–1864.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9), 2964–2996.

Degli Esposti, M., Spreckelsen, T., Gasparrini, A., Wiebe, D. J., Bonander, C., Yakubovich, A. R., & Humphreys, D. K. (2020). Can synthetic controls improve causal inference in interrupted time series evaluations of public health interventions?. International Journal of Epidemiology, 49(6), 2010–2020.

Bibliography

Graves, A., & Graves, A. (2012). Long short-term memory. Supervised Sequence Labelling with Recurrent Neural Networks, 37–45.

Grupp, T., Mishra, P., Reynaert, M., & Benthem, A. A. van. (2023). An evaluation of protected area policies in the european union.

Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. Arxiv Preprint Arxiv:2312.00752.

Miller, S., Johnson, N., & Wherry, L. R. (2019). Medicaid and mortality: new evidence from linked survey and administrative data.

Murphy, K. P. (2022). Probabilistic machine learning: an introduction. MIT press.

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.

Bibliography

Ryan, A. M., Krinsky, S., Kontopantelis, E., & Doran, T. (2016). Long-term evidence for the effect of pay-for-performance in primary care on mortality in the UK: a population study. The Lancet, 388(10041), 268–274.

Sabuncu, E., David, J., Bernède-Bauduin, C., Pépin, S., Leroy, M., Boëlle, P.-Y., Watier, L., & Guillemot, D. (2009). Significant reduction of antibiotic use in the community after a nationwide campaign in France, 2002–2007. Plos Medicine, 6(6), e1000084.

Schaffer, A. L., Dobbins, T. A., & Pearson, S.-A. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. BMC Medical Research Methodology, 21, 1–12.

Sheridan, P., McElroy, S., Casey, J., & Benmarhnia, T. (2022). Using the generalized synthetic control method to estimate the impact of extreme weather events on population health. Epidemiology, 33(6), 788–796.

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

Bibliography

Stechemesser, A., Koch, N., Mark, E., Dilger, E., Klösel, P., Menicacci, L., Nachtigall, D., Pretis, F., Ritter, N., Schwarz, M., & others. (2024). Climate policies that achieved major emission reductions: Global evidence from two decades. Science, 385(6711), 884–892.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37–45.

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