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.

Setup: event studies

Estimation of the effect of a treatment when data is:

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

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.

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.

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.

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 studies 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

- Reminder on difference-in-differences
- Synthetic control method: balancing method (similar to propensity score weighting)
- Conditional DID: doubly robust method combining outcomes and treatment models
- Methods without controls: if we have time

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

- Reminder on difference-in-differences
- Synthetic control method: balancing method (similar to propensity score weighting)
- Conditional DID: doubly robust method combining outcomes and treatment models
- Methods without controls: if we have time

Table of contents

Reminder on difference-in-differences

Difference-in-differences

History

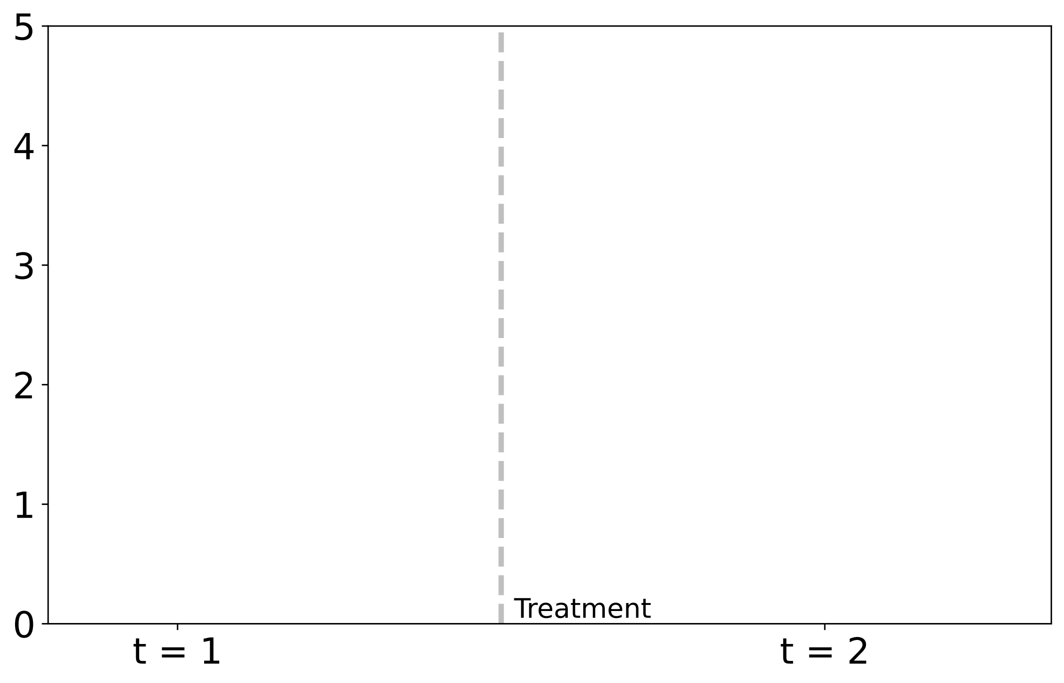
- First documented example (though not formalized): John Snow showing how cholera spread through the water in London (Snow, 1855)[1]
- Modern usage introduced formally by (Ashenfelter, 1978), applied to labor economics

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

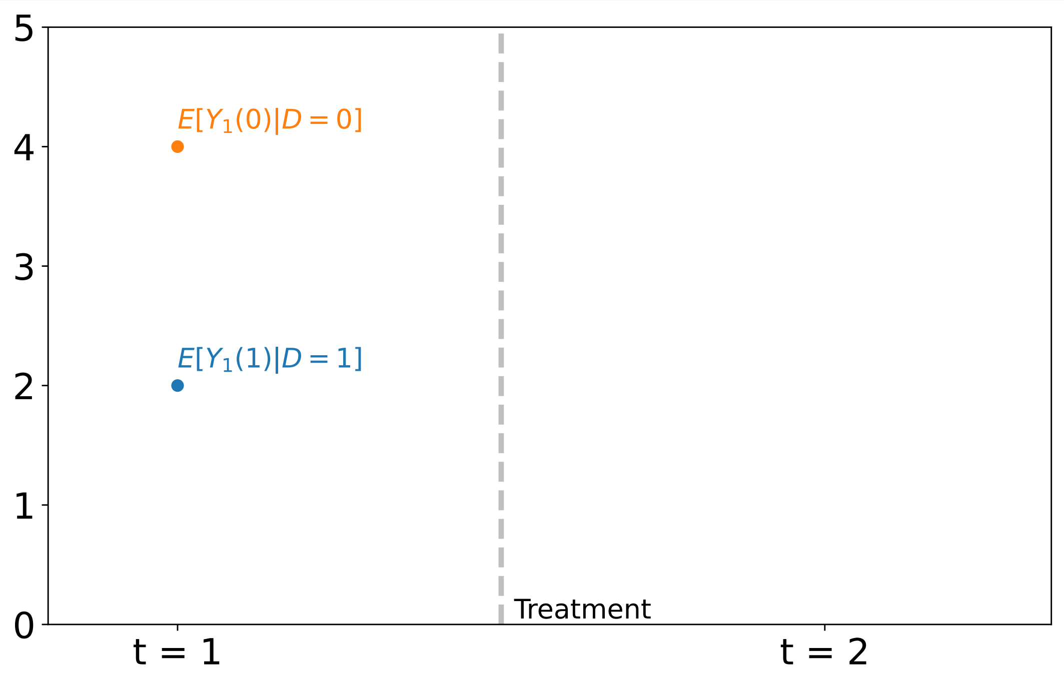
[1] Good description: https://mixtape.scunning.com/09-difference_in_differences#john-snows-cholera-hypothesis

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



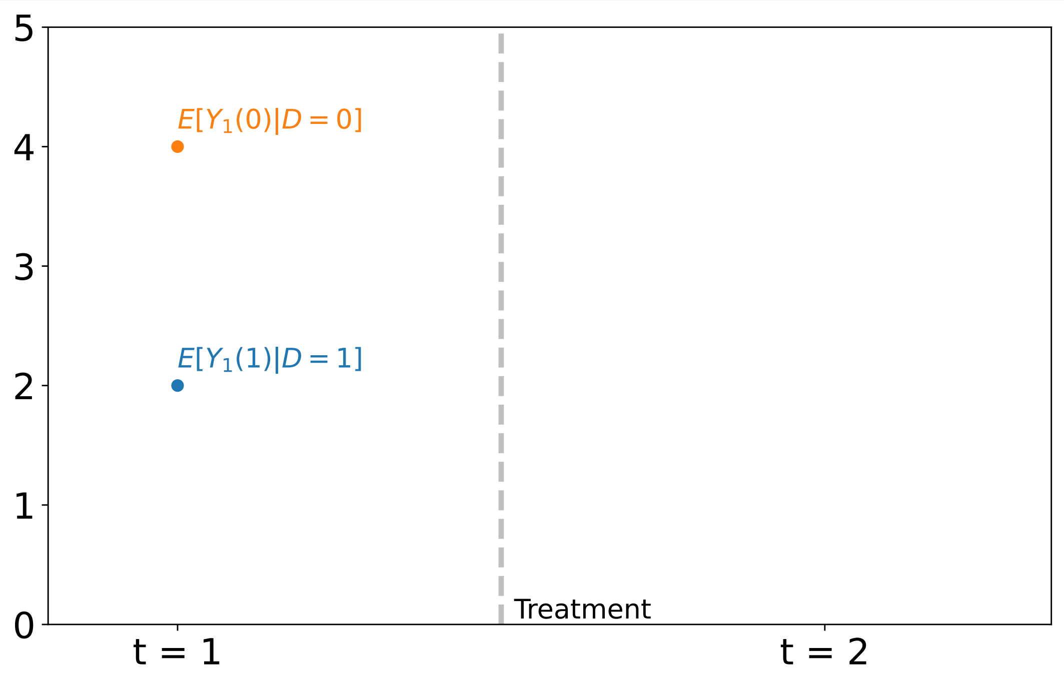
Difference-in-differences framework

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



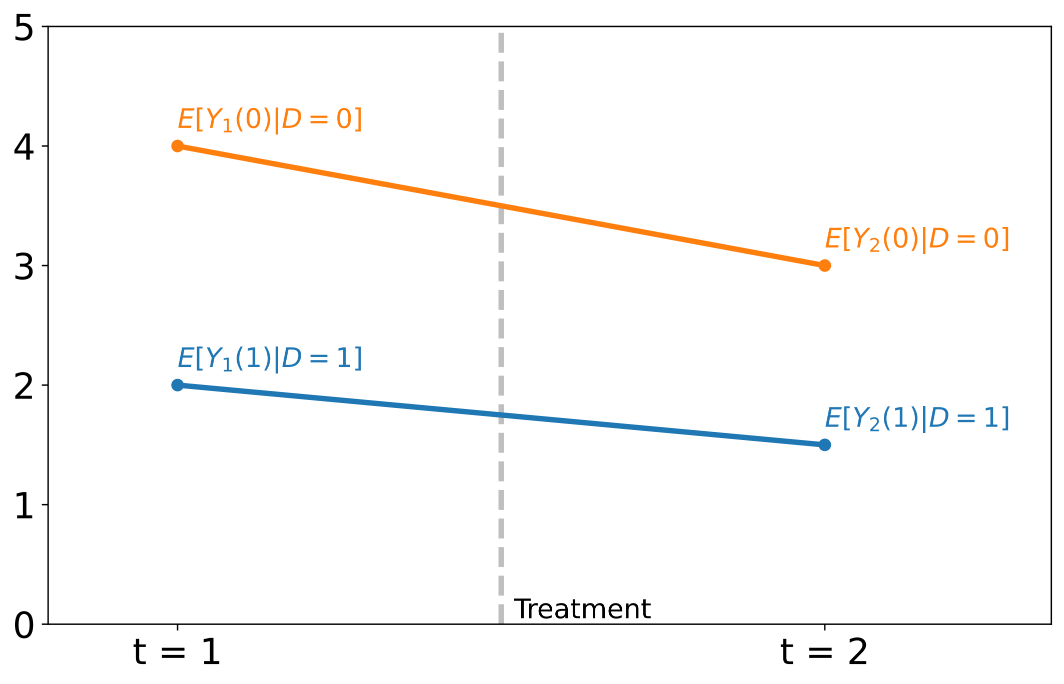
Difference-in-differences framework

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

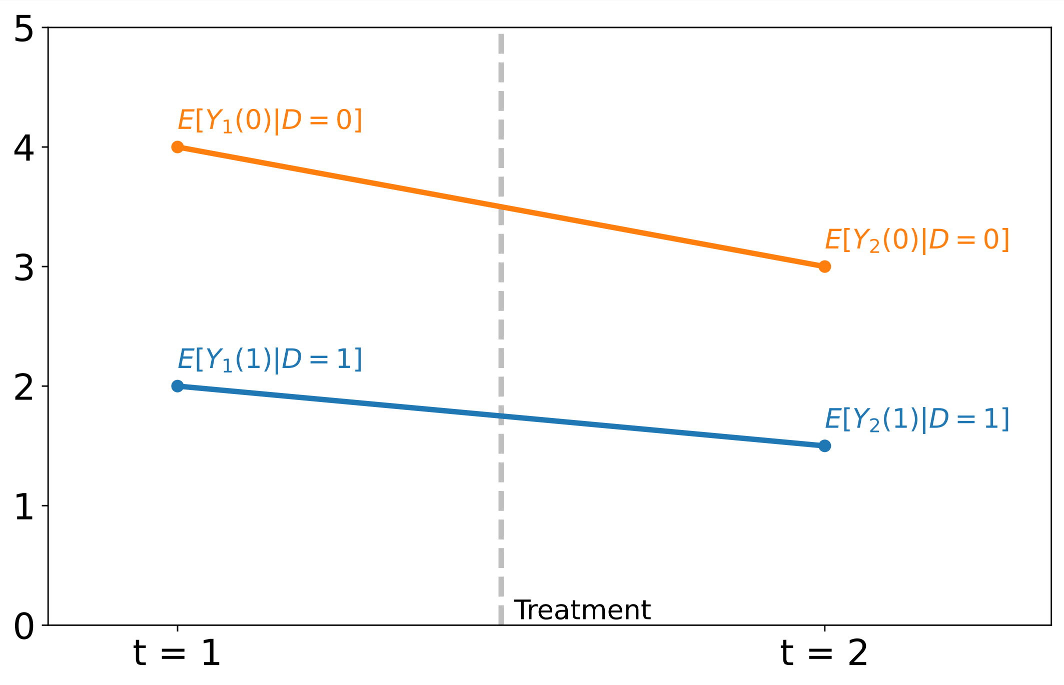
Our target is the average treatment effect on the treated (ATT)



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

Difference-in-differences framework

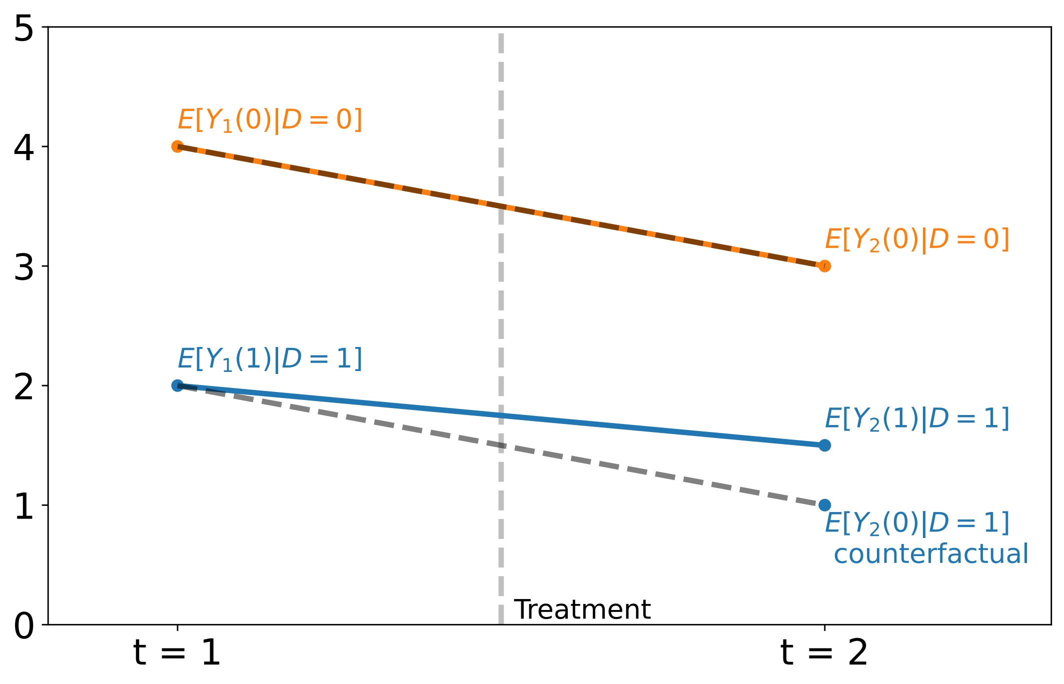
Our target is the average treatment effect on the treated (ATT)



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

First assumption, parallel trends

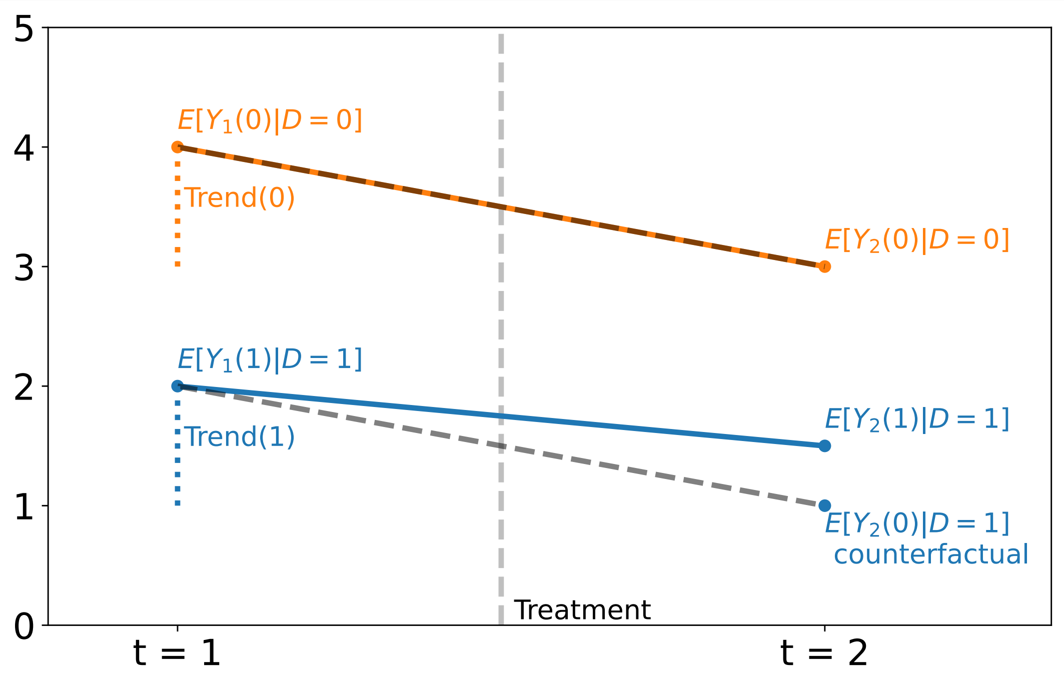
$$\mathbb{E}[Y_2(0) - Y_1(0) \mid D = 1] = \mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]$$



Difference-in-differences framework

First assumption, parallel trends

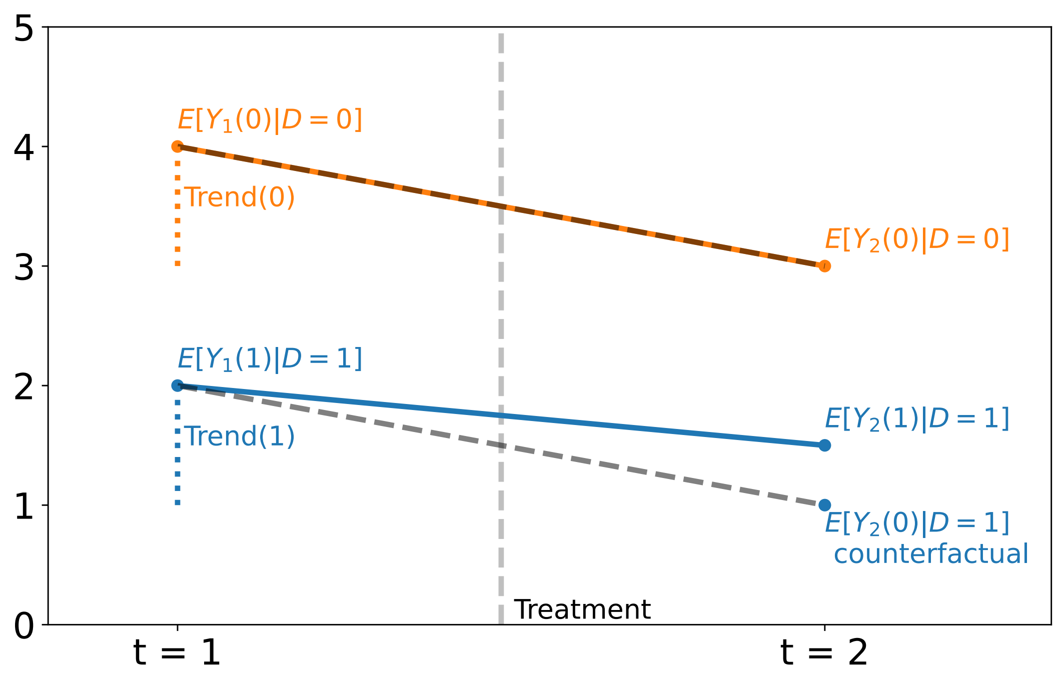
$$\underbrace{[Y_2(0) - Y_1(0) \mid D = 1]}_{\color{blue}{\mathrm{Trend}(1)}} = \underbrace{\mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]}_{\color{orange}{\mathrm{Trend}(0)}}$$



Difference-in-differences framework

First assumption, parallel trends

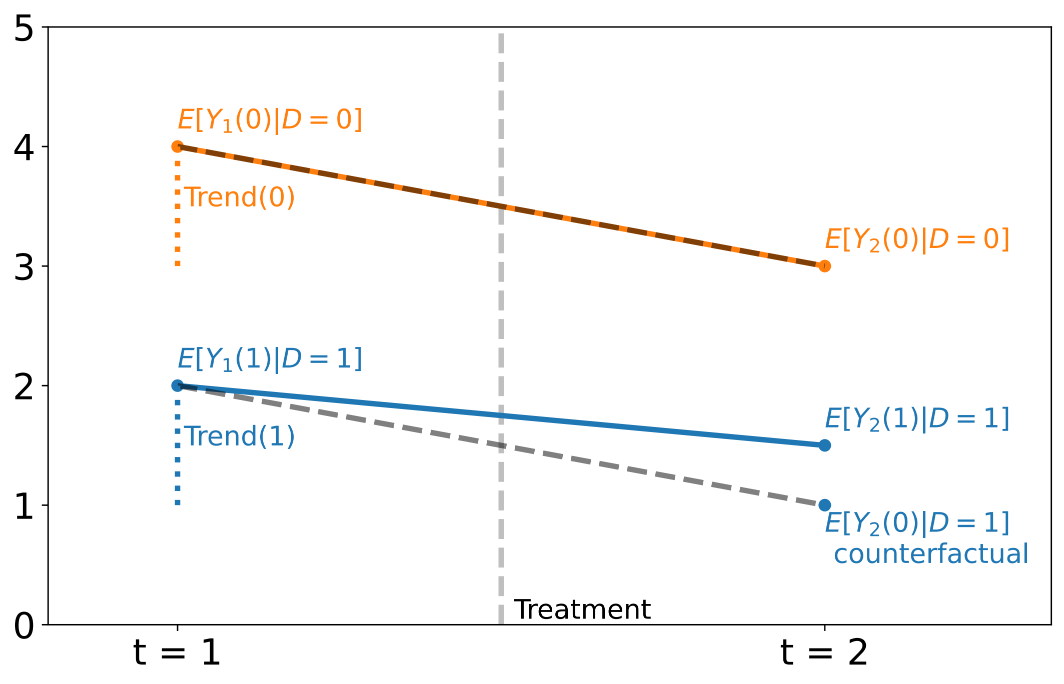
$$\mathbb{E}[Y_2(0) \mid D = 1] = \mathbb{E}[Y_1(0) \mid D = 1] + \mathbb{E}[Y_2(0) - Y_1(0) \mid D = 0]$$



Difference-in-differences framework

First assumption, parallel trends

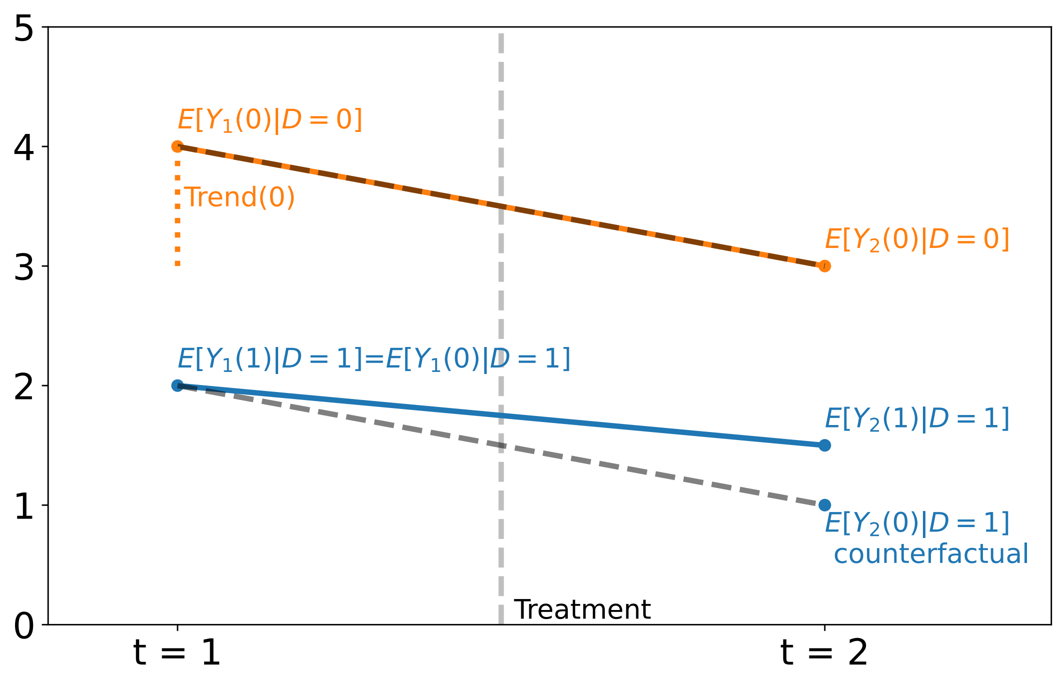
$$\mathbb{E}[Y_2(0) \mid D=1] = \underbrace{[Y_1(0) \mid D=1]}_{\text{unobserved counterfactual}} + \mathbb{E}[Y_2(0) - Y_1(0) \mid D=0]$$



Difference-in-differences framework

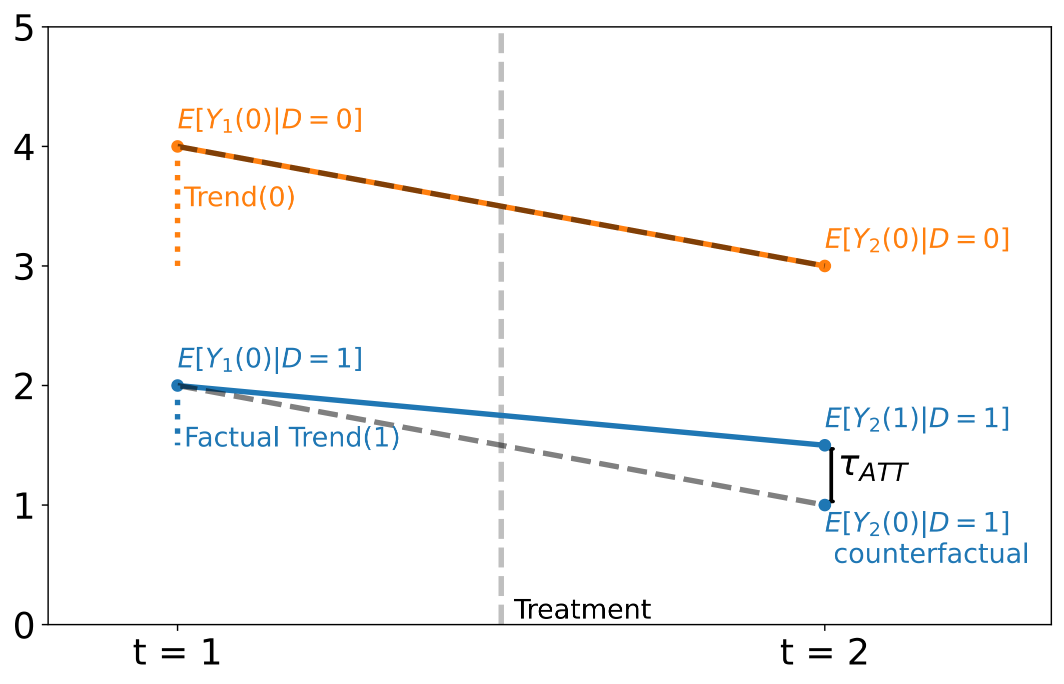
Second assumption, no anticipation of the treatment

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



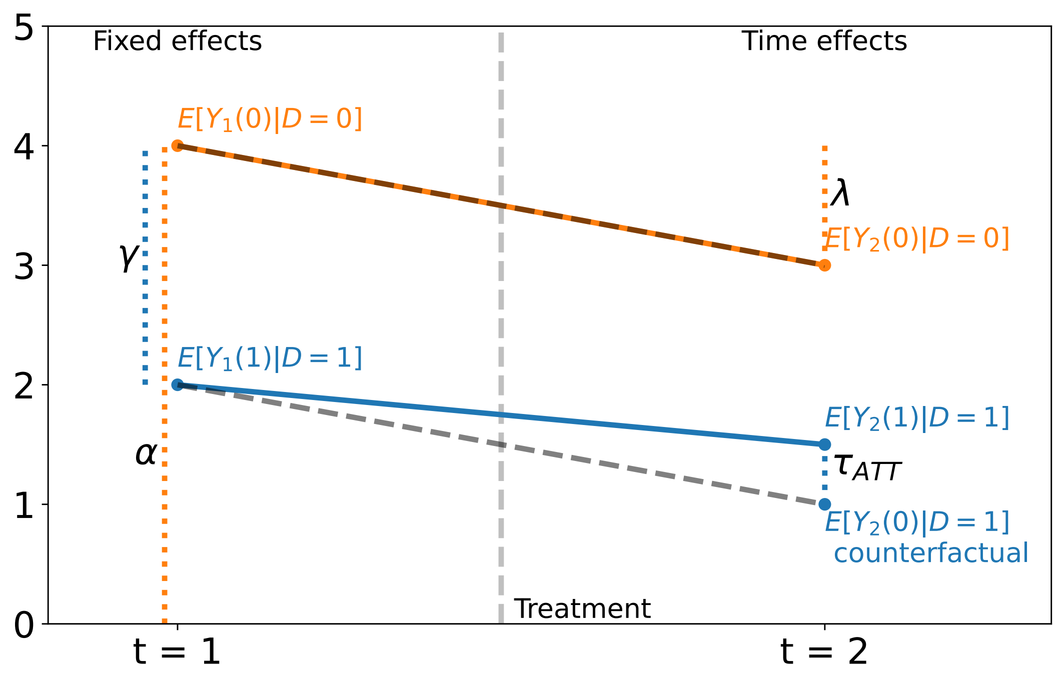
Difference-in-differences framework: identification of ATT

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



Estimation: link with two way fixed effect (TWFE)

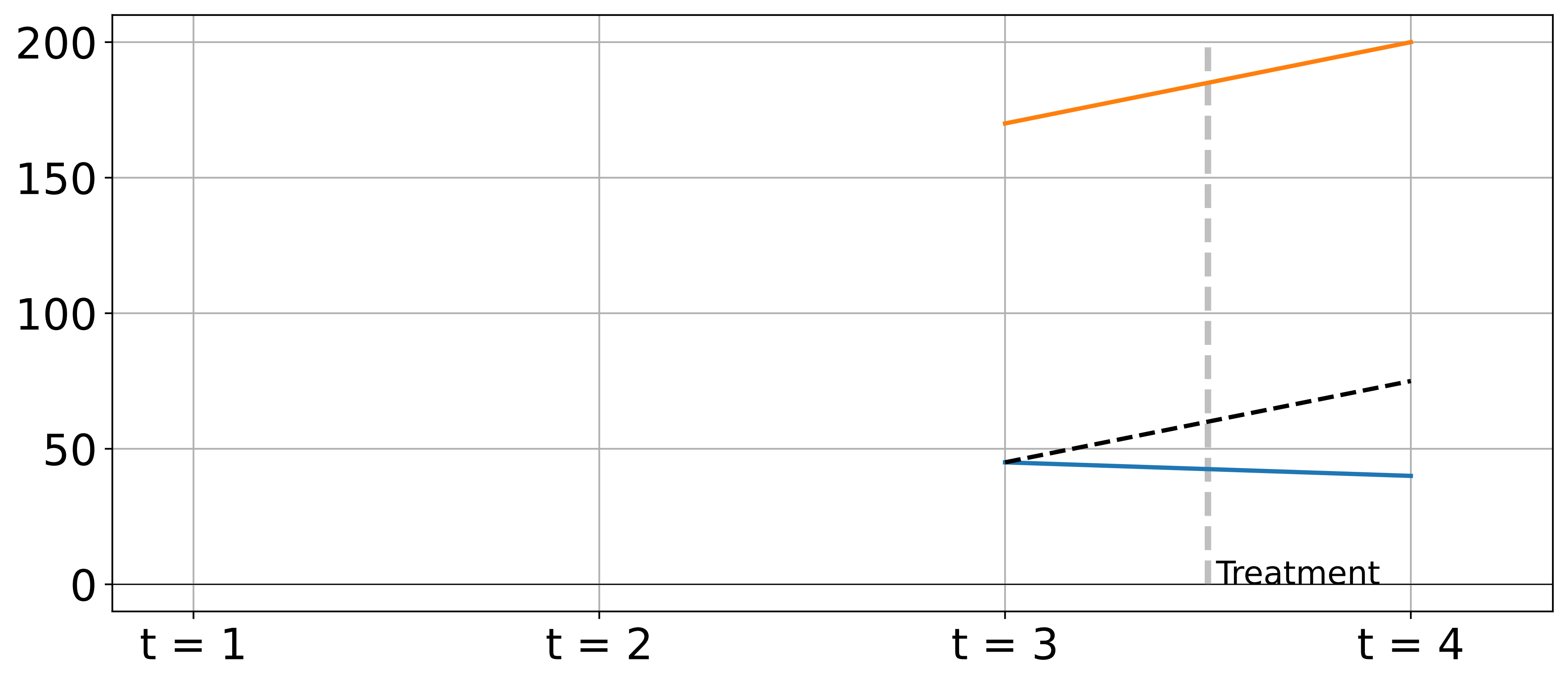
$$Y = \alpha + \gamma D + \lambda \mathbb{1}(t=2) + \tau_{\text{ATT}} D \mathbb{1}(t=2)$$



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

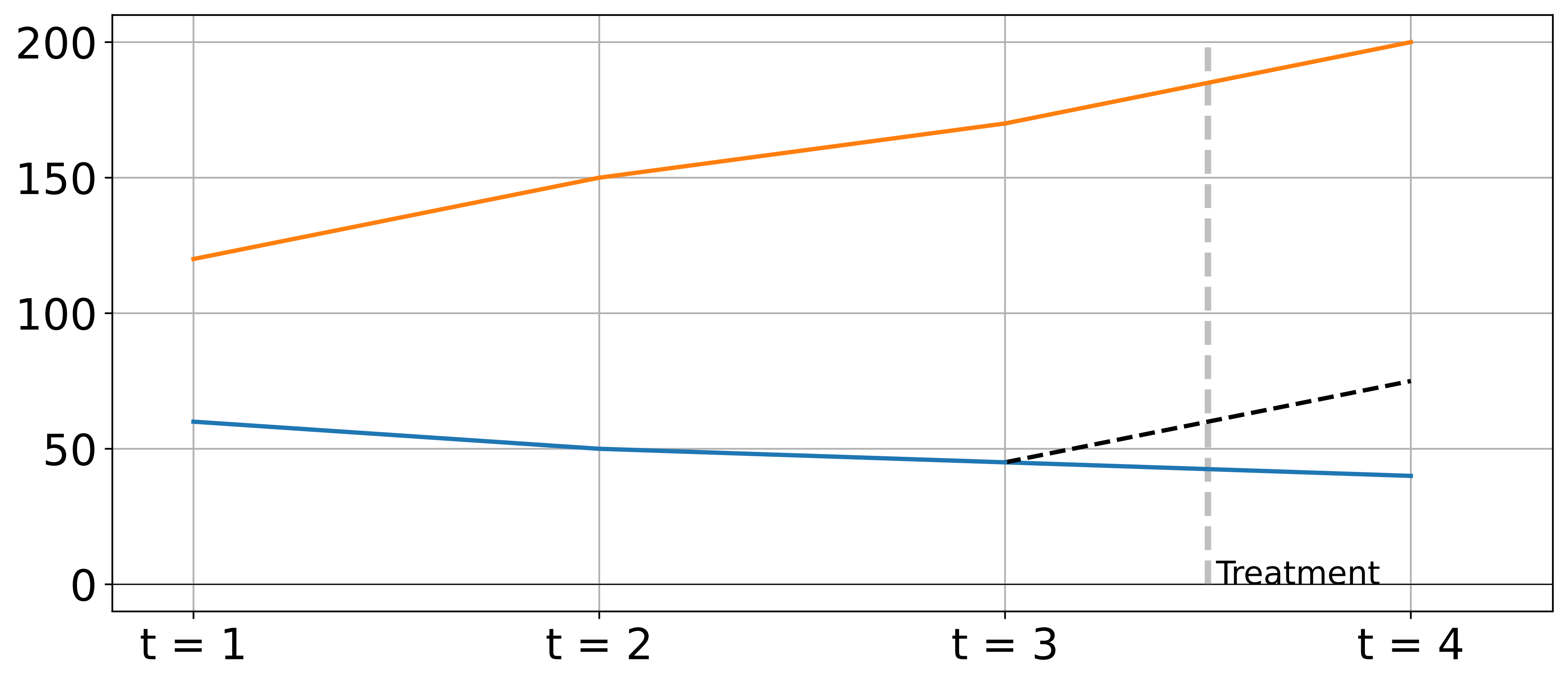
Failure of the parallel trend assumption

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

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

- Very 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

- Very 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),

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)

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)

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

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

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

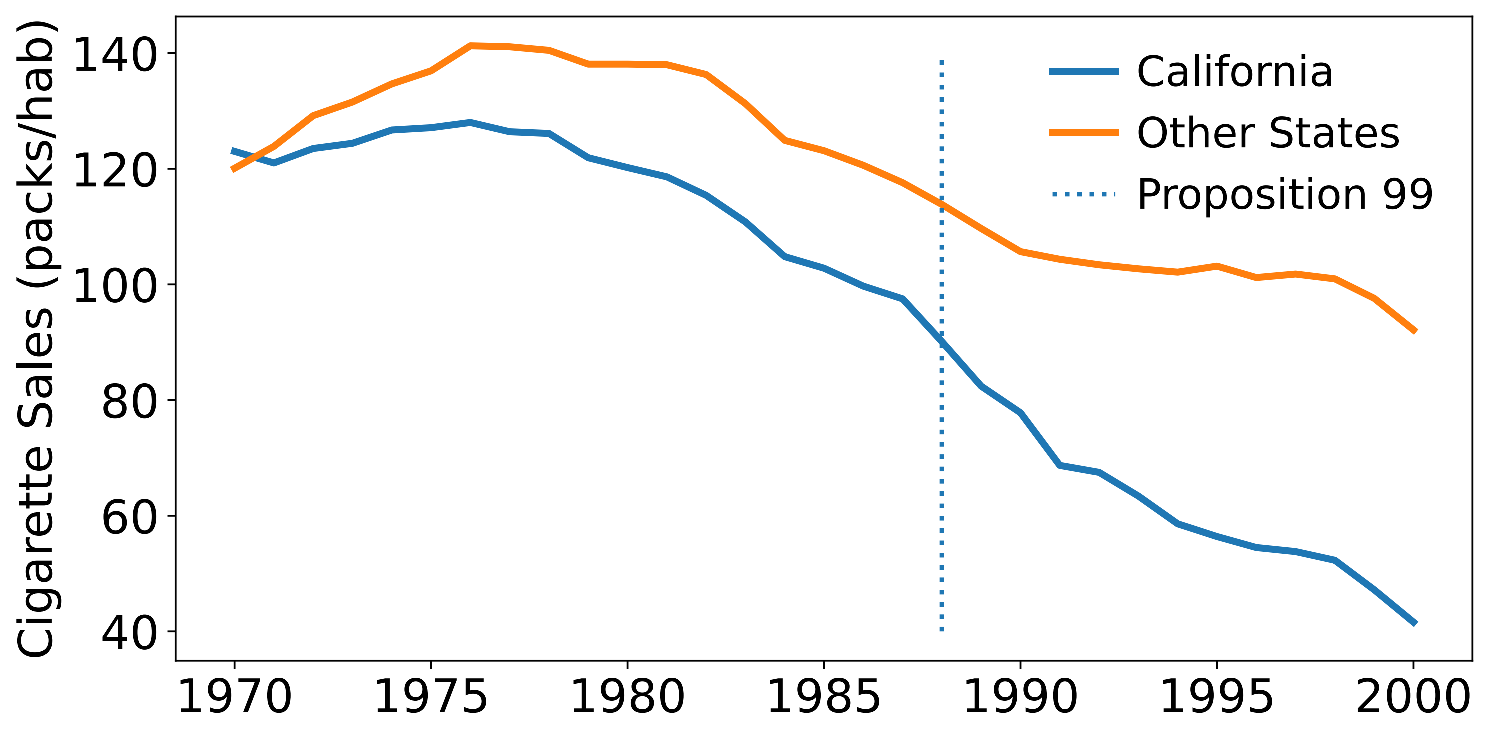
Treated unit, $j = 1$: California as from 1988

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

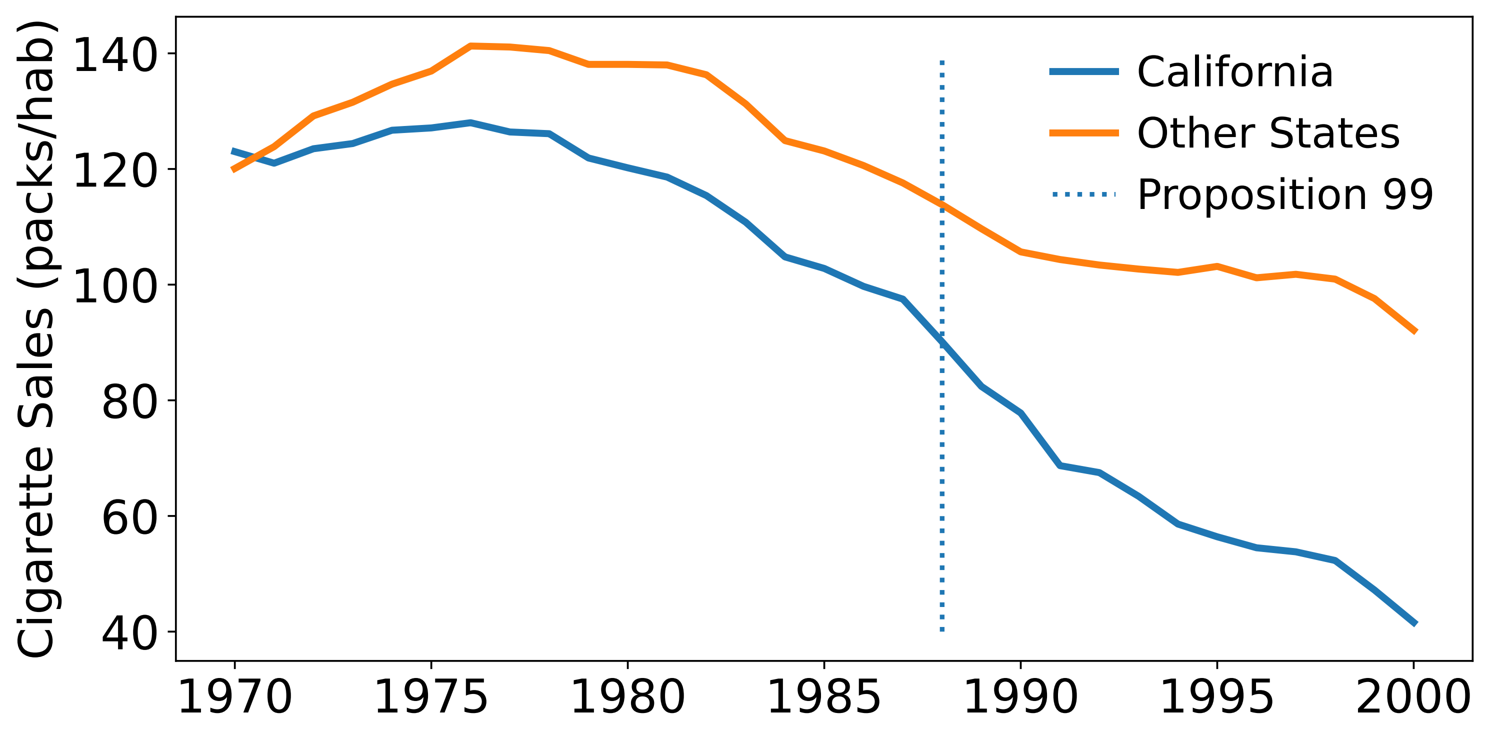
Time period: $t \in \{1, ..T\} = \{1970, ..2000\}$ **and treatment time** $T_0 = 1988$

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

Synthetic control example: plot the data

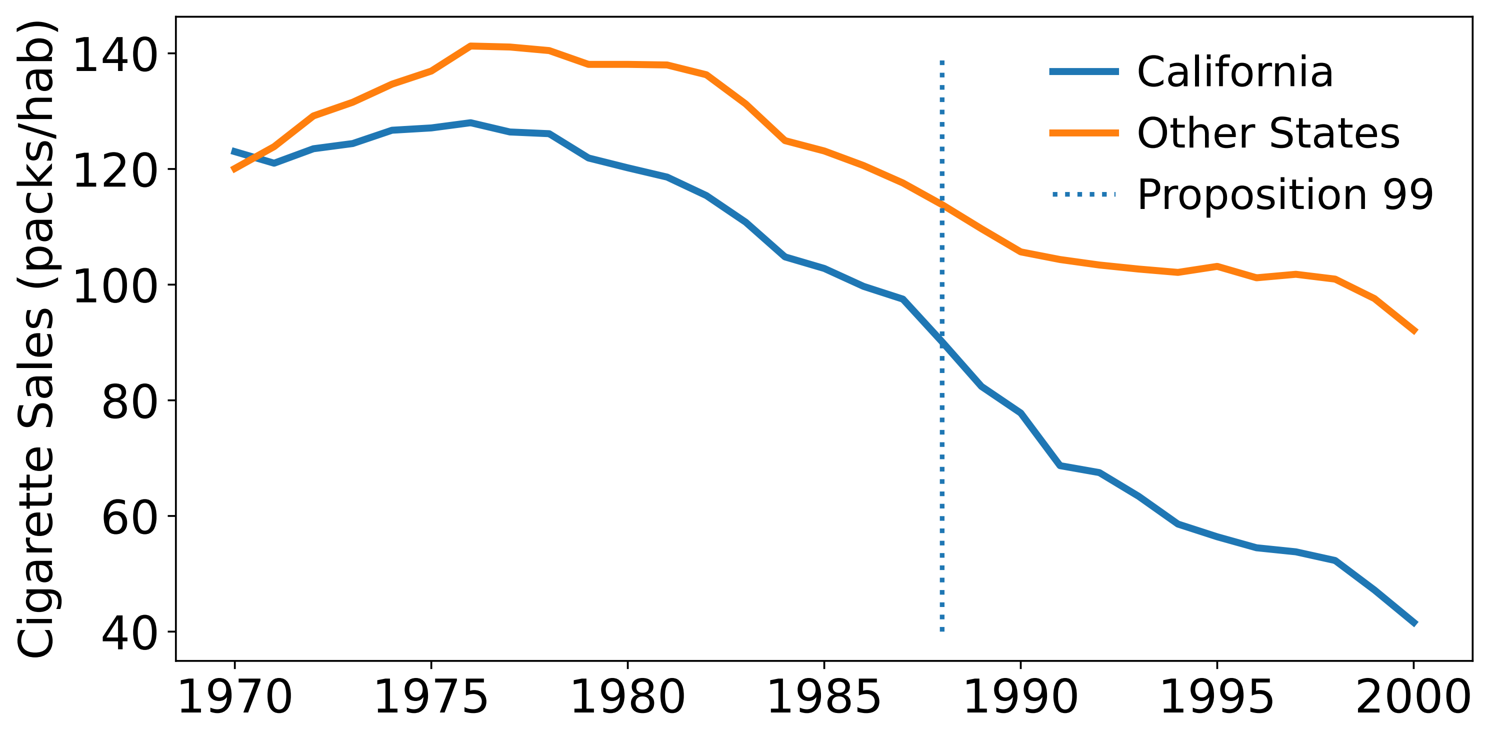


Synthetic control example: plot the data



😯 Decrease in cigarette sales in California.

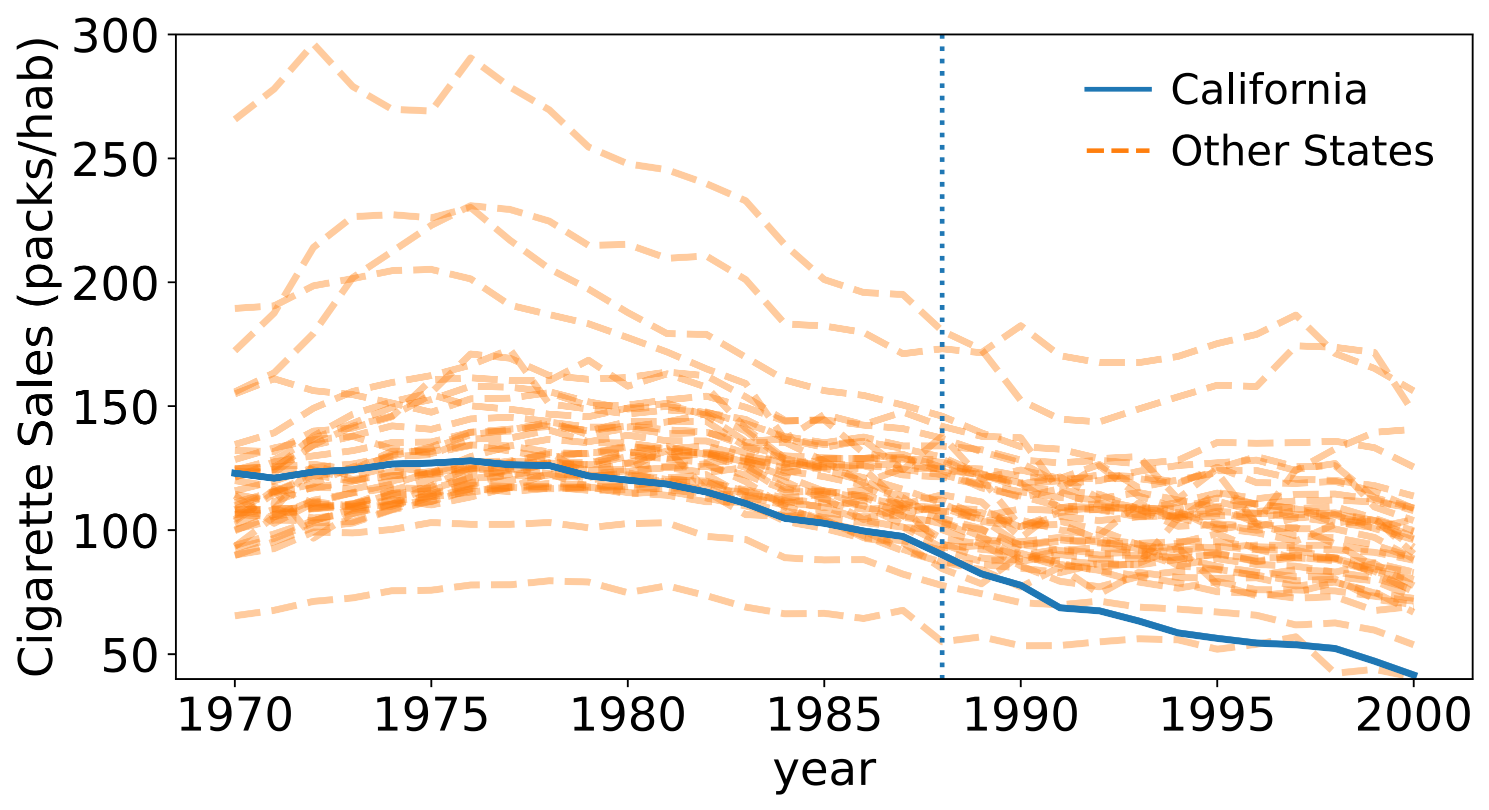
Synthetic control example: plot the data



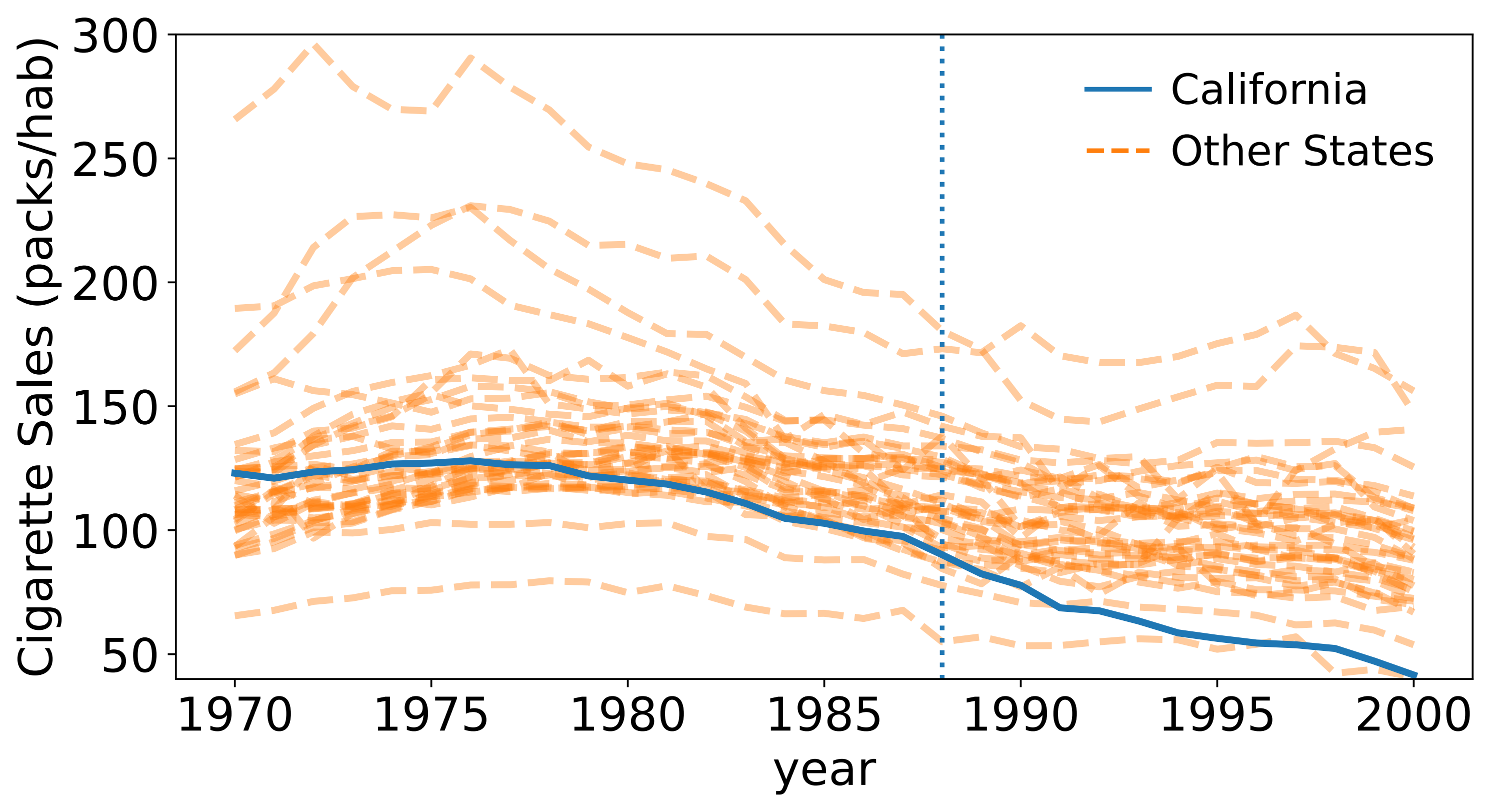
😯 Decrease in cigarette sales in California.

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

Synthetic control example: plot the data



Synthetic control example: plot the data

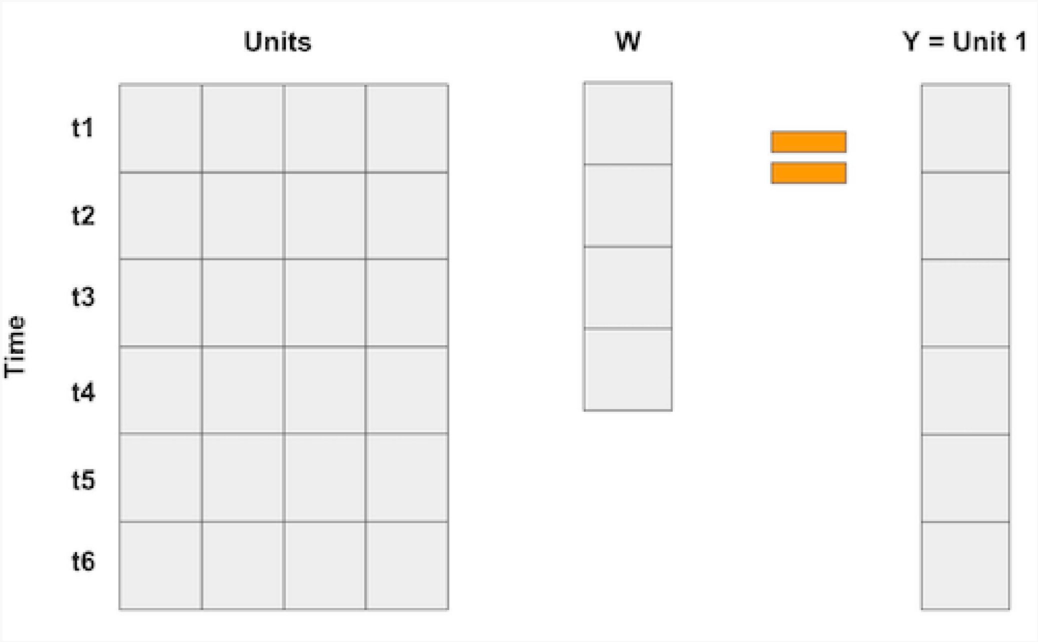


💡 Force parallel trends: Find a weighted average of other states that predicts well the pre-treatment trend of California (before $T_0 = 1988$).

Synthetic control as weighted average of control outcomes

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

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



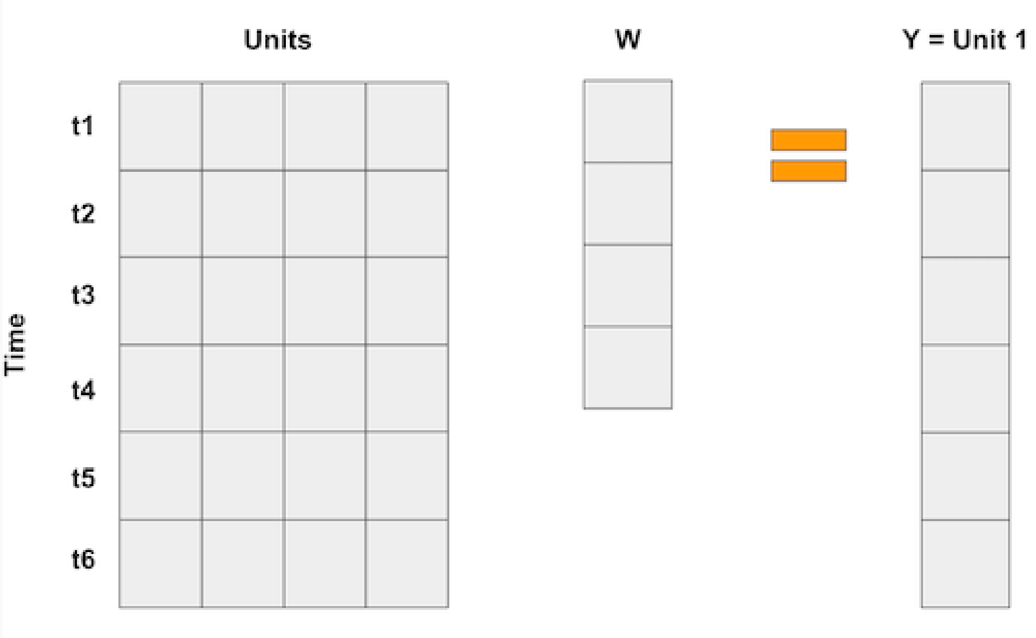
Synthetic control as weighted average of control outcomes

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.



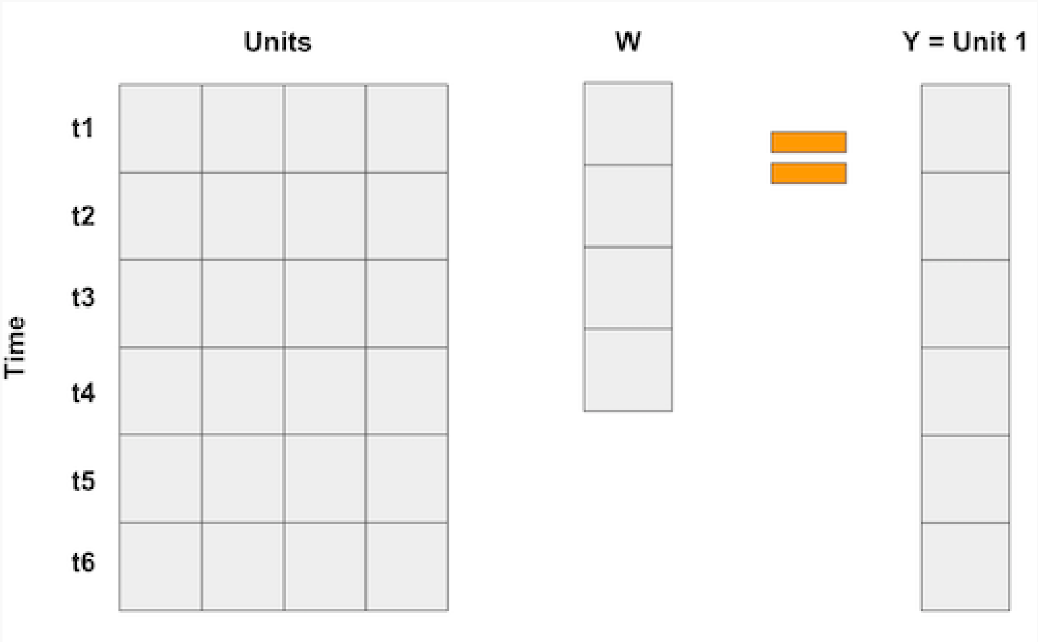
Synthetic control as weighted average of control outcomes

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

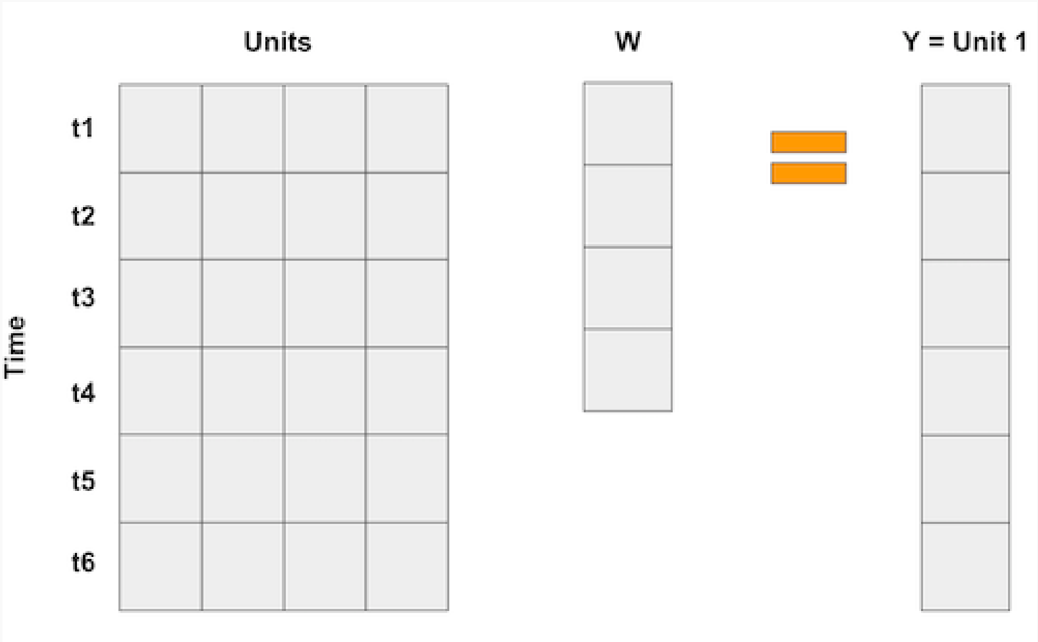
Build a predictor for $Y_{1,t}$ (California):

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

🤔 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 (Wager, 2024, chapter 7)



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

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

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

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

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

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: $X_{\text{control}} = \left(X_2, .., X_{n_0+1}\right) \in R^{p\times n_0}$

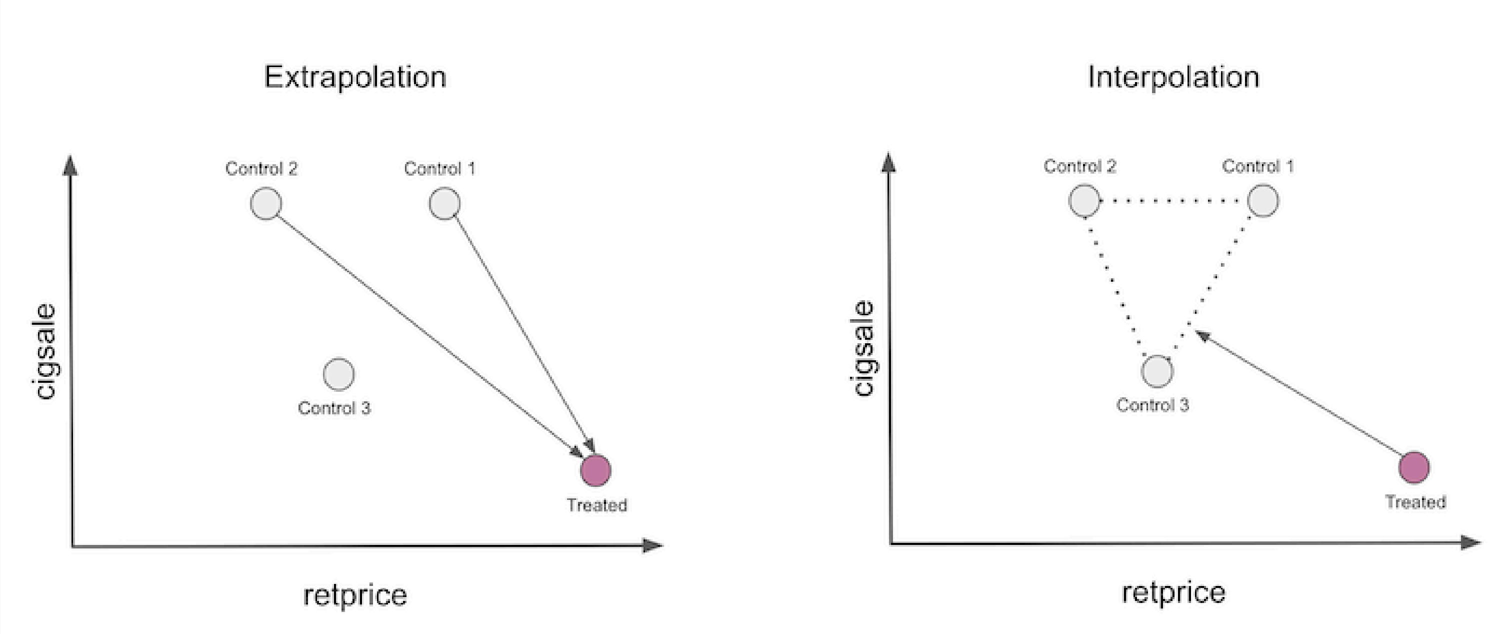
Minimization problem with constraints

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

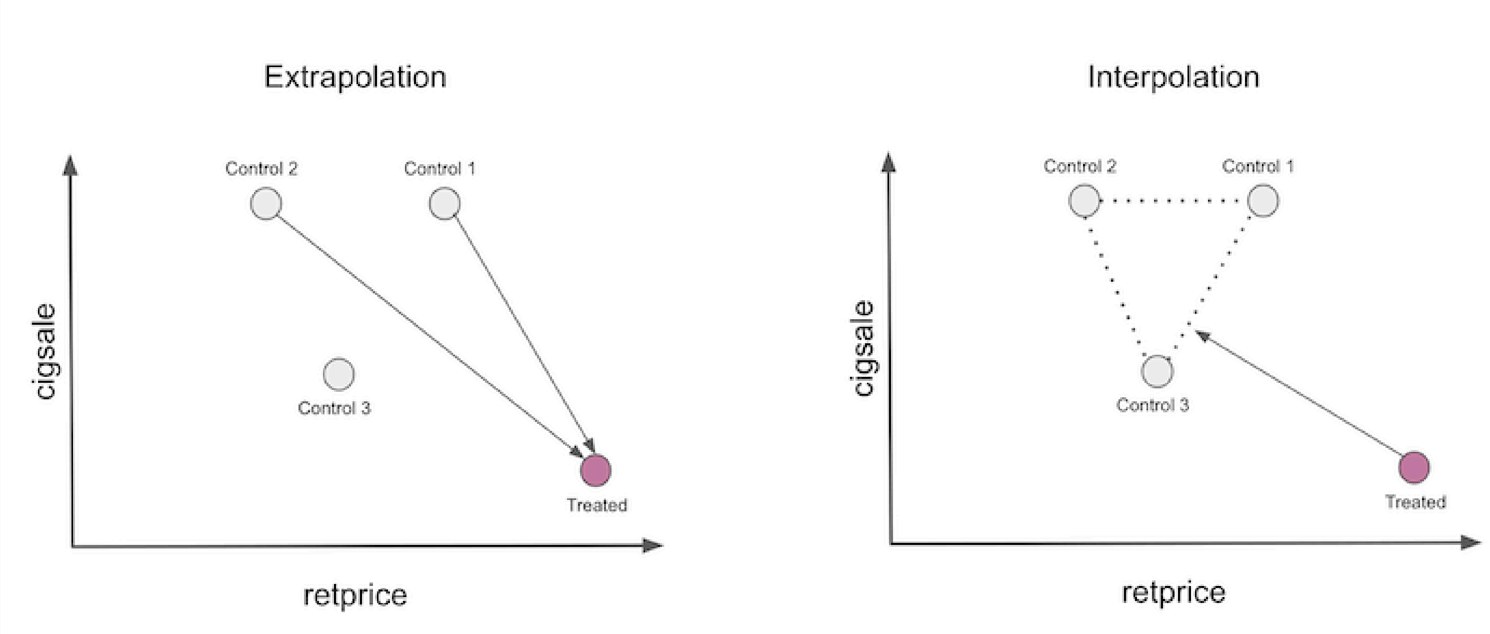
$$s.t. \; w_j \geq 0,$$

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

This is called interpolation (vs extrapolation)



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 (both are *optimization with constraints on a simplex*).

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

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

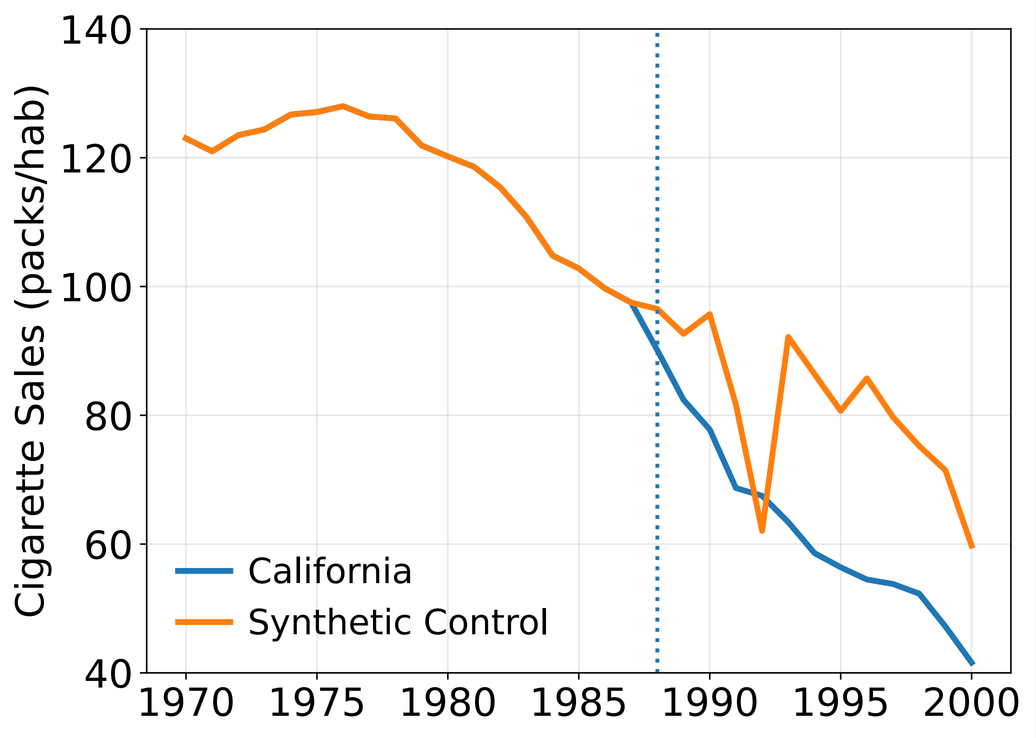
$p = 2T_0$ covariates:

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

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



Synthetic controls: Extrapolation failure with unconstrained weight

$p = 2T_0$ covariates:

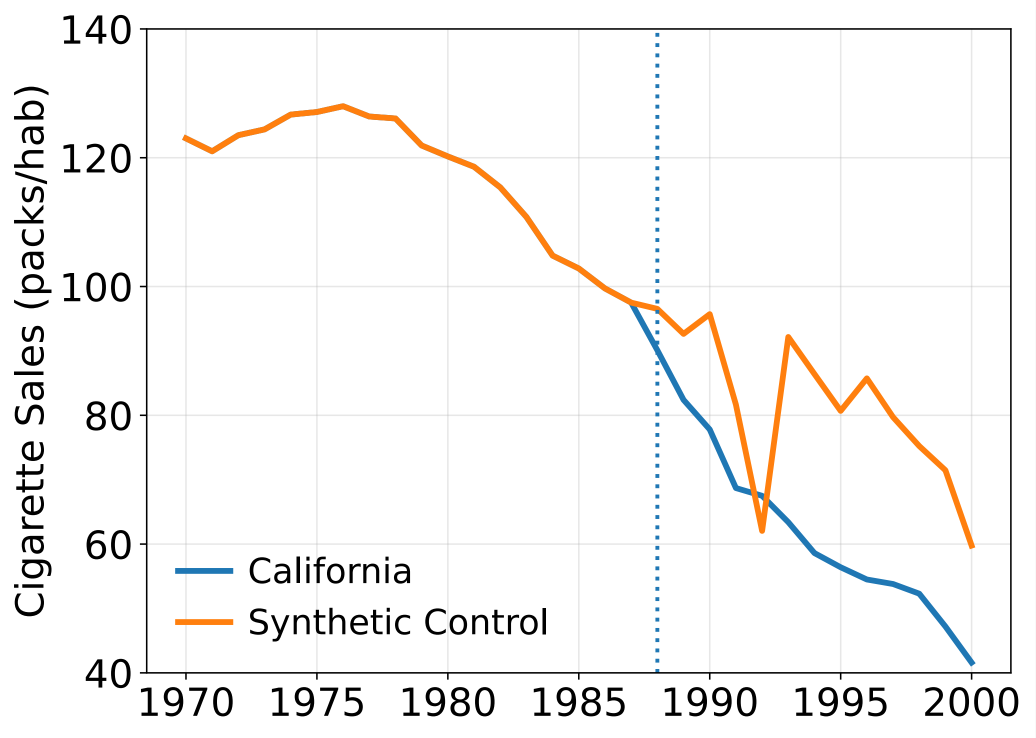
$$X_j = \begin{pmatrix} Y_{j,1} \\ .. \\ Y_{j,T_0} \\ Z_{j,1} \\ .. \\ Z_{j,T_0} \end{pmatrix}^T \in R^{2T_0}$$

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

😭 **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:
   $V = \text{diag}\left(\text{var}\left(Y_{j,1}\right)^{-1}, .., \text{var}\left(Y_{j,T_0}\right)^{-1}, \text{var}\left(Z_{j,1}\right)^{-1}, .., \text{var}\left(Z_{j,T_0}\right)^{-1}\right)$.
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:
   $V = \mathrm{diag}\left(\mathrm{var}\left(Y_{j,1}\right)^{-1}, .., \mathrm{var}\left(Y_{j,T_0}\right)^{-1}, \mathrm{var}\left(Z_{j,1}\right)^{-1}, .., \mathrm{var}\left(Z_{j,T_0}\right)^{-1}\right)$.
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:
   $V = \operatorname{diag}\left(\operatorname{var}\left(Y_{j,1}\right)^{-1}, .., \operatorname{var}\left(Y_{j,T_0}\right)^{-1}, \operatorname{var}\left(Z_{j,1}\right)^{-1}, .., \operatorname{var}\left(Z_{j,T_0}\right)^{-1}\right)$.
3. Minimize the pre-treatment mean squared prediction error (MSPE) of the treated unit:

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

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

This solution is solved by running two optimization problems:

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

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, $\|X\|_V = \|X\|_2$.

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

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

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

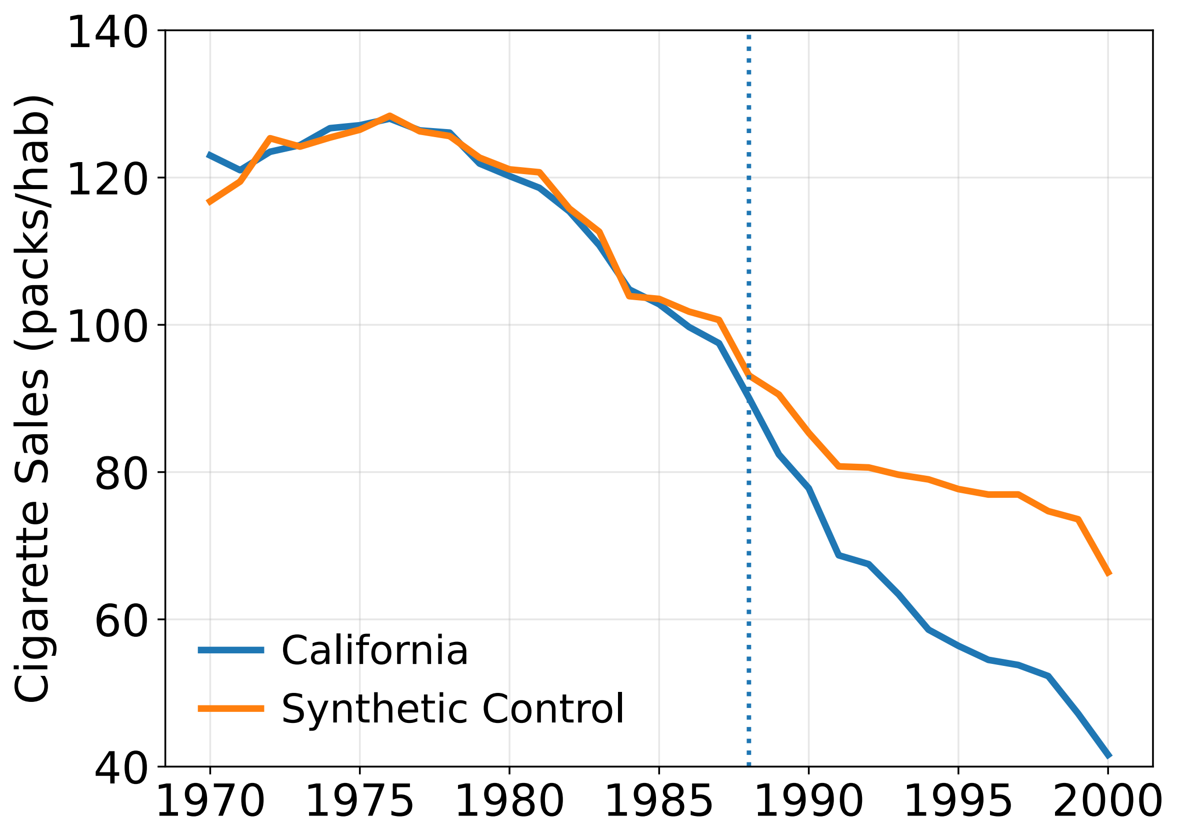
Synthetic controls: estimation without the outer optimization problem

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

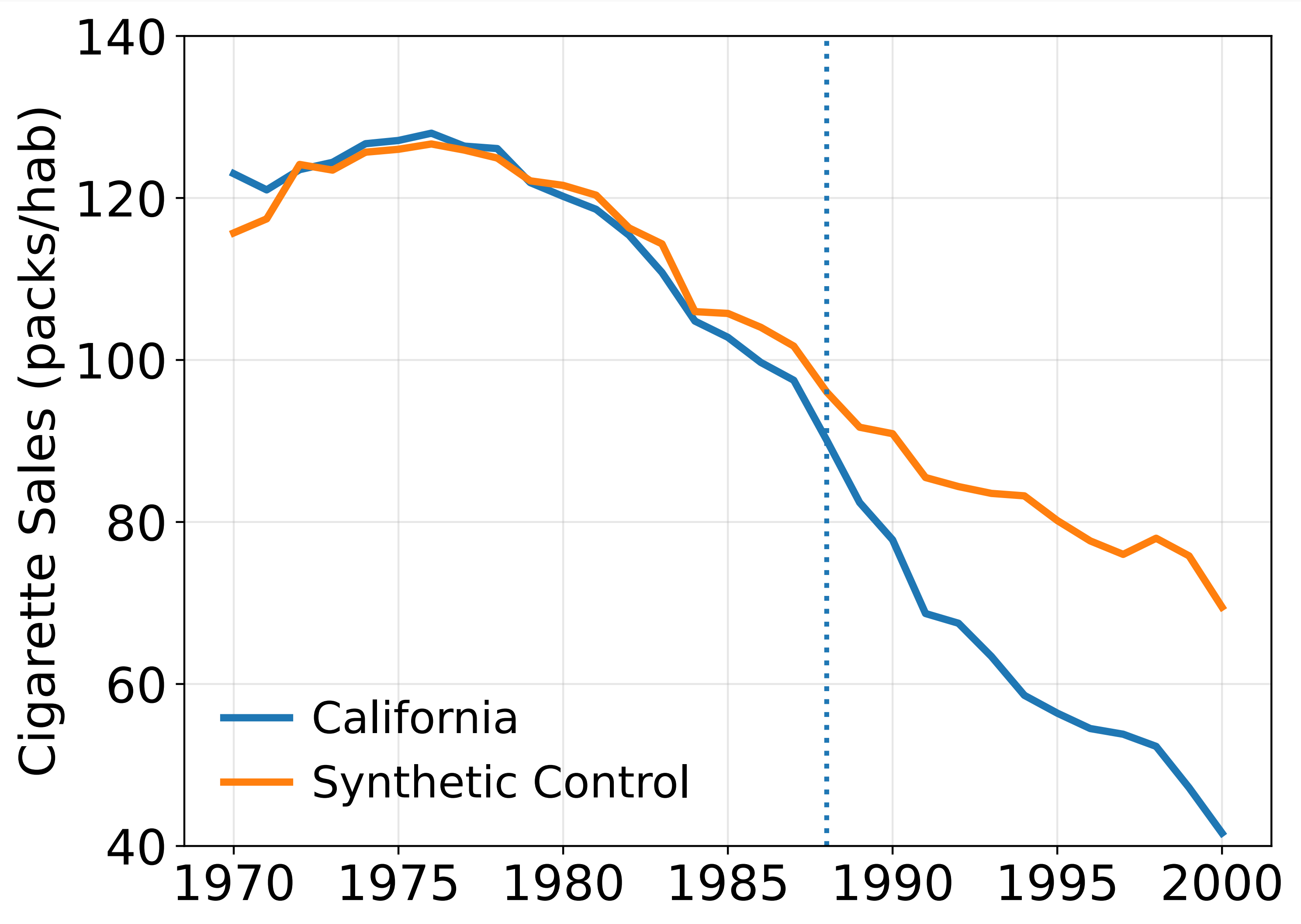
Y cigarette sales, Z cigarette prices.

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.

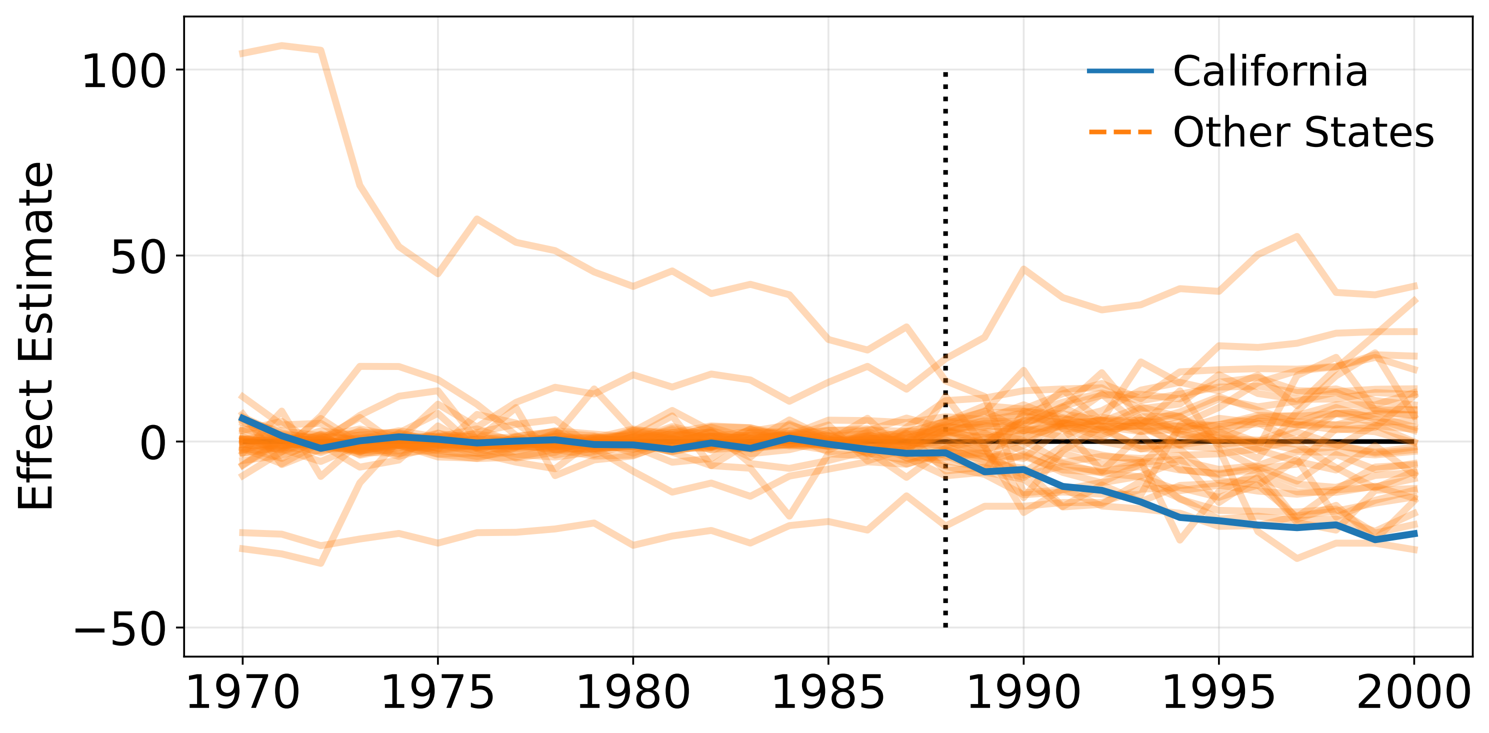
Synthetic controls: inference with Placebo tests

Idea of 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.

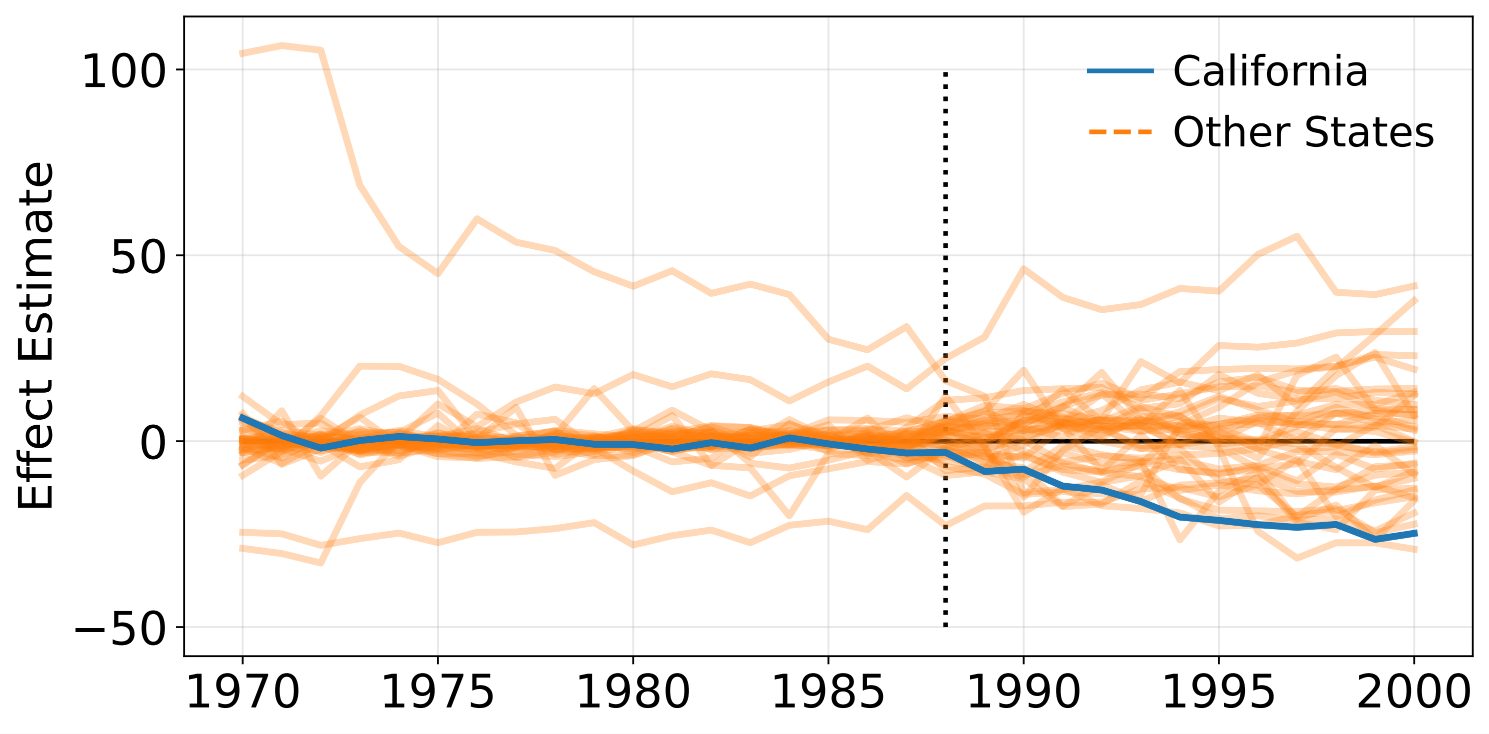
Synthetic controls: inference with Placebo tests, example

Placebo estimation for all 38 control states



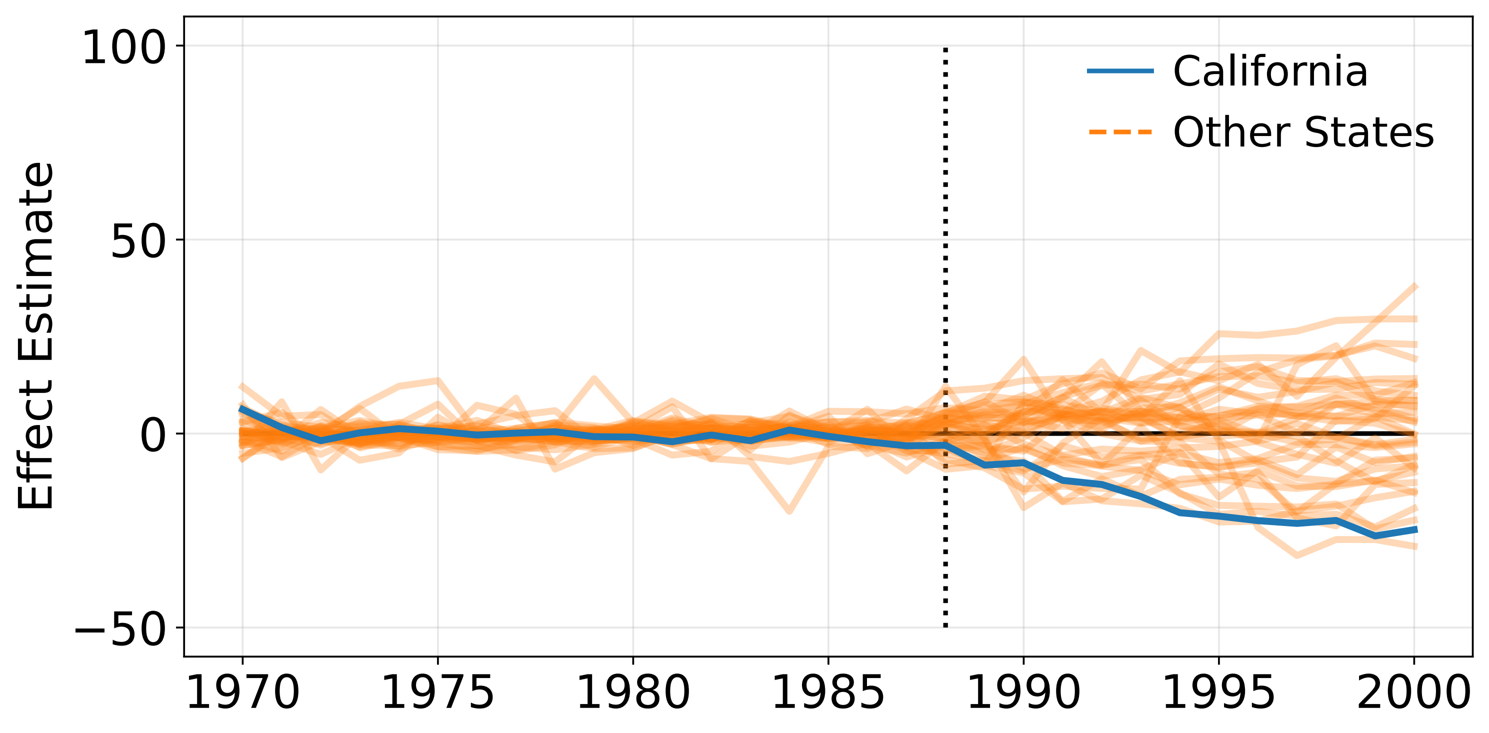
Synthetic controls: inference with Placebo tests, example

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.

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

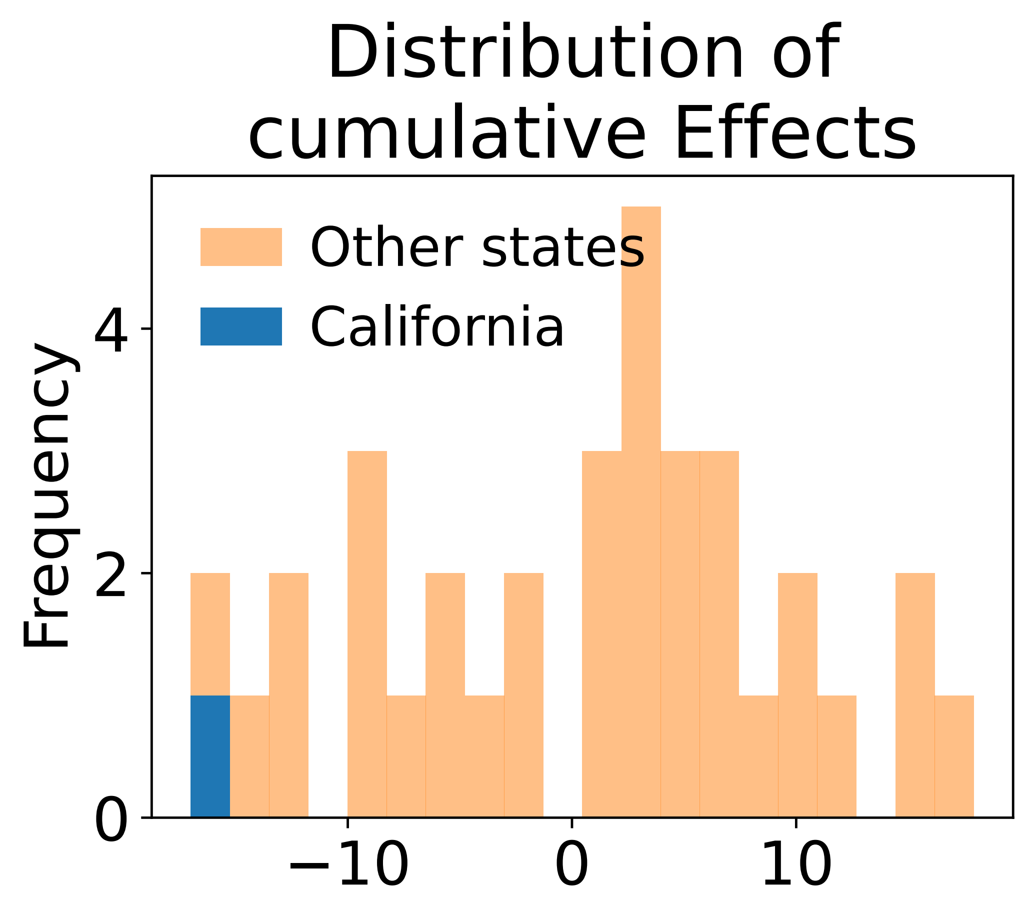


I removed the states above the 90 percentiles of the distribution of the pre-treatment fit.

Synthetic controls: inference with Placebo tests, example

California absolute cumulative effect

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

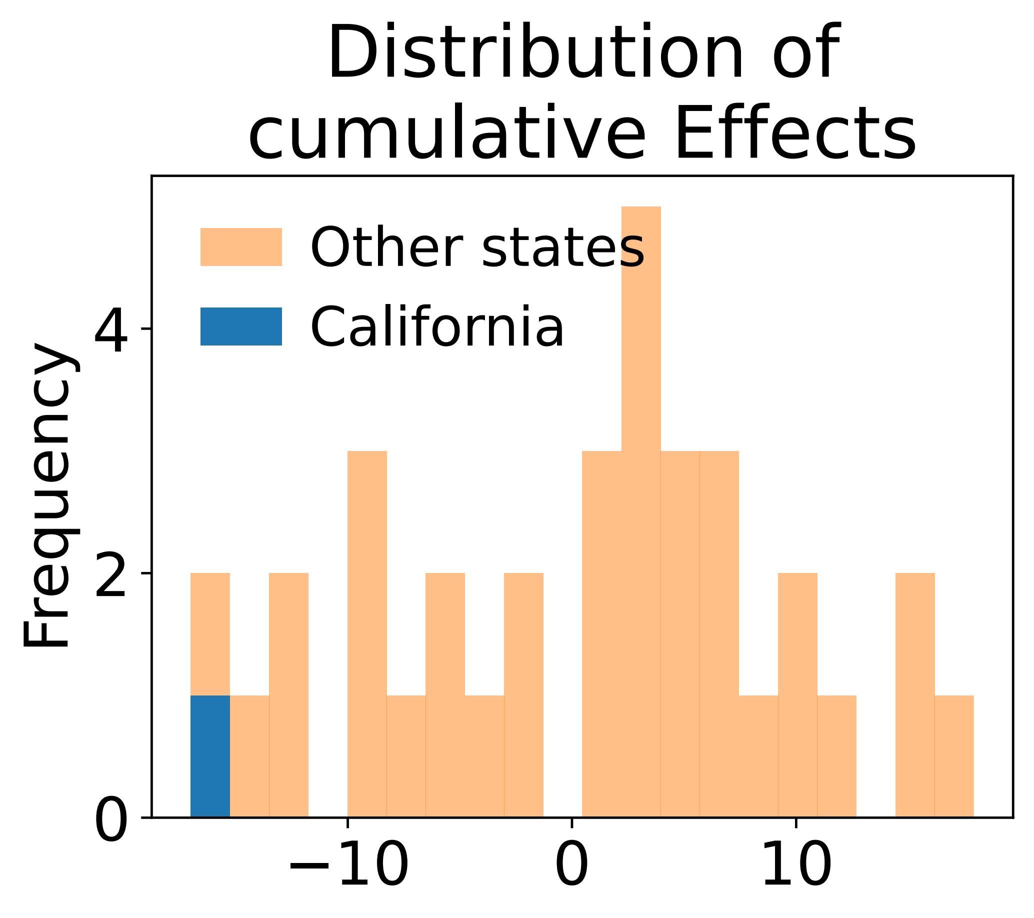


California absolute cumulative effect

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

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: inference with conformal prediction

Synthetic controls: Take-away

Pros

- More convincing for parallel trends assumption.
- Simple for multiple time periods.
- Gives confidence intervals.

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. See (Arkhangelsky et al., 2021) for discussions.
- Still requires the no-anticipation assumption.

Conditional difference-in-differences

Time-series modelisation: methods without a control group

Interrupted Time Series

Idea

- Compare the evolution of the outcome before and after the treatment
- The treatment effect is the difference between the two trends

Example

-

State space models

Take-away

Good references for event studies

- The causal mixtape: https://mixtape.scunning.com/09-difference_in_differences
- Causal inference for the brave and true: https://matheusfacure.github.io/python-causality-handbook/13-Difference-in-Differences.html

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.

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

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

Bibliography

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

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.

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

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.