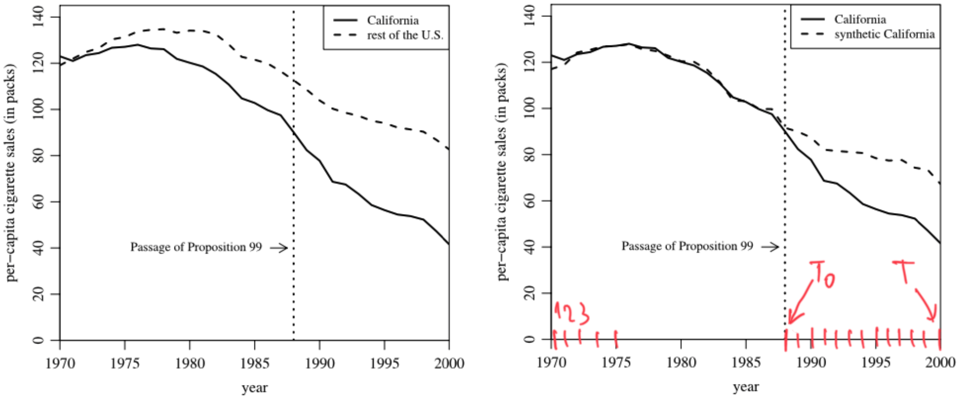
Synthetic controls method

Lukáš Lafférs

Matej Bel University, Dept. of Mathematics

- In many situations the treatment happens on an aggregate level (city, state).
- We may not have a natural unit to use as a control, the treated unit is simply very different than the rest
- We create it artificially (hence synthetic) by weighting other units so that the characteristics of the weighted unit resembles the one of the treated unit

Example: Tabacco control program and cigarettes sales

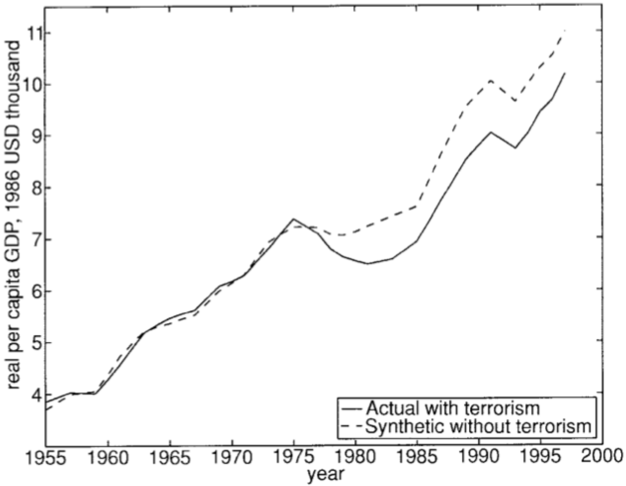


Abadie, Diamond and Hainmueller (2010)

It's about comparison

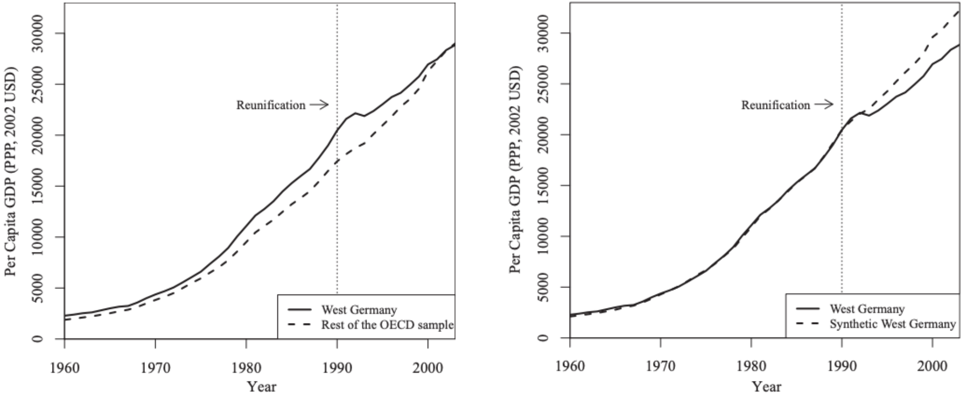
> *The synthetic control method is based on the idea that, when the units of observation are a small number of aggregate entities, a combination of unaffected units often provides a more appropriate comparison than any single unaffected unit alone.*
>
> (Abadie, 2021, p. 393)

Example: The economic cost of a conflict



Abadie and Gardezabal (2003)

Example: Reunification of Germany and Economic growth



Abadie, Diamond and Hainmueller (2015)

- time $1, 2, \ldots, T$
- $J+1$ units, 1 is treated in $T_0+1, \ldots, T$
- Synthetic control is a weighted average of the J control units. $(w_2, \ldots, w_{J+1})$ with $w_j \geq 0, \sum_{j=2}^{J+1} w_j = 1$
- Weights w_j^* are chosen optimally to make the synthetic control similar to the control one in observed characteristics.
- Synthetic control estimator is

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* \cdot Y_{jt}$$

Choosing the weights

What does optimally mean?

We need some metric. Assume k variables $X_1, \ldots, X_k$. E.g. we can choose weighted Euclidean metric.

Pre-intervention outcomes are also included in the set of predictors!

$$\arg\min_{w} \sum_{h=1}^{k} \left(X_{h1} - \sum_{j=2}^{J+1} w_h \cdot X_{hj} \right)^2$$

Choosing the weights

What does optimally mean?

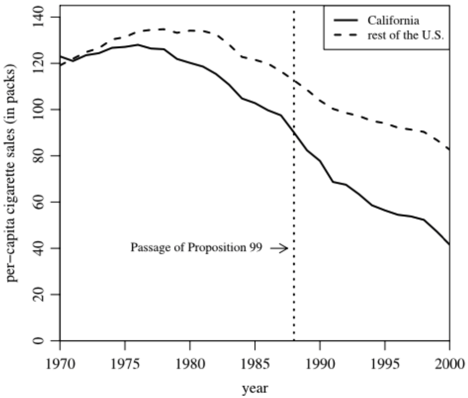
We need some metric. Assume k variables $X_1, \dots, X_k$. E.g. we can choose weighted Euclidean metric.

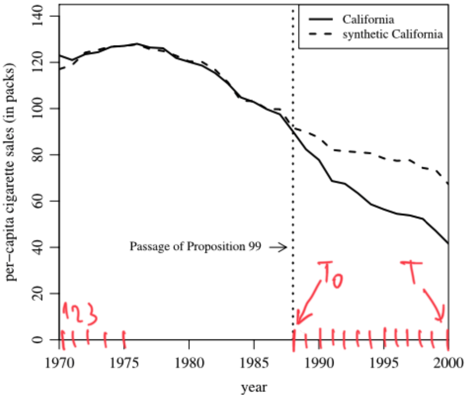
Pre-intervention outcomes are also included in the set of predictors!

Larger weights v_h on more important predictors.

$$\arg\min_{w} \sum_{h=1}^{k} v_h \cdot \left(X_{h1} - \sum_{j=2}^{J+1} w_h \cdot X_{hj} \right)^2$$

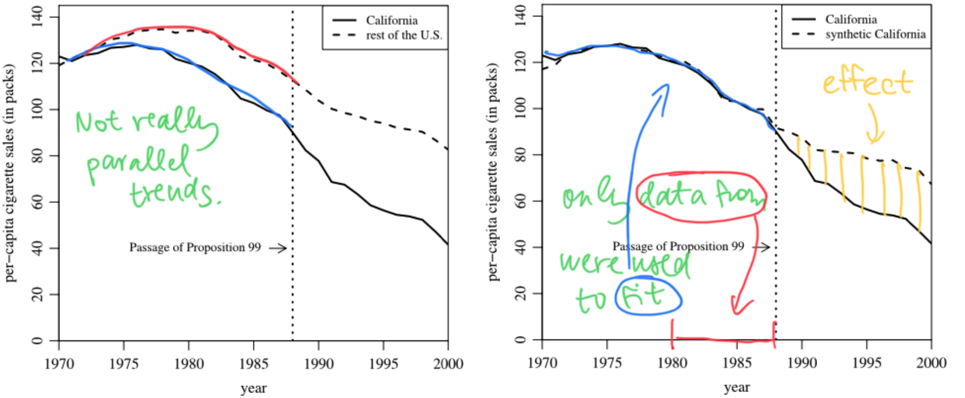
Example: Tabacco again





Abadie, Diamond and Hainmueller (2010)

Example: Tabacco again



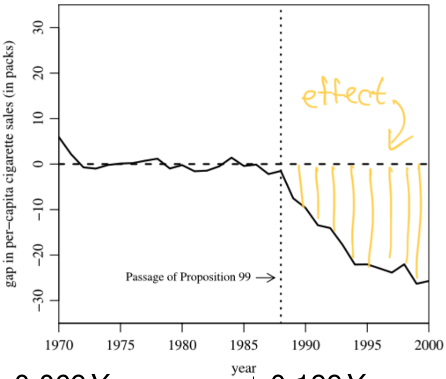
Abadie, Diamond and Hainmueller (2010)

Weights

Table 2. State weights in the synthetic California

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	–	Nebraska	0
Arizona	–	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	–
Connecticut	0.069	New Mexico	0
Delaware	0	New York	–
District of Columbia	–	North Carolina	0
Florida	–	North Dakota	0
Georgia	0	Ohio	0
Hawaii	–	Oklahoma	0
Idaho	0	Oregon	–
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	–	Vermont	0
Massachusetts	–	Virginia	0
Michigan	–	Washington	–
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

sparse (many zeros)



$$Y_{synth,t} = 0.164\,Y_{Colorado,t} + 0.069\,Y_{Connecticut,t} + 0.199\,Y_{Montana,t} + 0.234\,Y_{Nevada,t} + 0.334\,Y_{Utah,t}$$

$$\underbrace{\hat{\tau}_{California,t}}_{\text{effect}} = \underbrace{Y_{California,t}}_{\text{real outcome}} - \underbrace{Y_{synth,t}}_{\text{synthetic control}}$$

Balance

Table 1. Cigarette sales predictor means

Variables	California Real	California Synthetic	Average of 38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15–24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

Abadie, Diamond and Hainmueller (2010)

Example: Economic cost of a conflict

	Basque Country (1)	Spain (2)	"Synthetic" Basque Country (3)
Real per capita GDP[a]	5,285.46	3,633.25	5,270.80
Investment ratio (percentage)[b]	24.65	21.79	21.58
Population density[c]	246.89	66.34	196.28
Sectoral shares (percentage)[d]			
Agriculture, forestry, and fishing	6.84	16.34	6.18
Energy and water	4.11	4.32	2.76
Industry	45.08	26.60	37.64
Construction and engineering	6.15	7.25	6.96
Marketable services	33.75	38.53	41.10
Nonmarketable services	4.07	6.97	5.37
Human capital (percentage)[e]			
Illiterates	3.32	11.66	7.65
Primary or without studies	85.97	80.15	82.33
High school	7.46	5.49	6.92
More than high school	3.26	2.70	3.10

Abadie and Gardeazabal (2003)

Example: Reunification of Germany

	West Germany (1)	Synthetic West Germany (2)	OECD average (3)	Austria (nearest neighbor) (4)
GDP per capita	15,808.9	15,802.2	13,669.4	14,817.0
Trade openness	56.8	56.9	59.8	74.6
Inflation rate	2.6	3.5	7.6	3.5
Industry share	34.5	34.4	33.8	35.5
Schooling	55.5	55.2	38.7	60.9
Investment rate	27.0	27.0	25.9	26.6

real X_1

synthetic $X_0 W^*$

average $\bar{X}_0$

Application from Abadie, Diamond and Hainmueller (2015), reprinted from Abadie (2021)

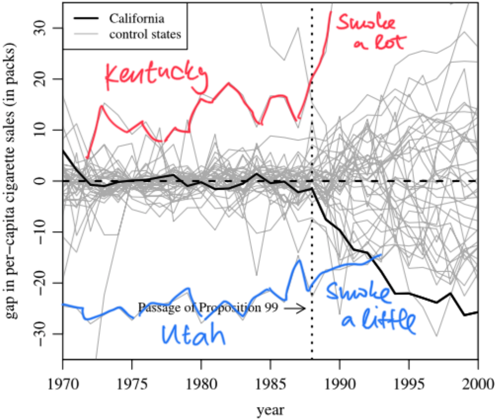
Statistical Inference

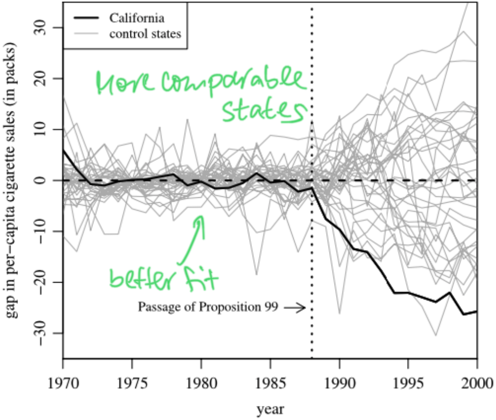
Use permutation method.

- Consider every control as a ”fake” treatment and estimate placebo effect
- Compare the effect for treated unit with those placebo effects
- Effect for the treated should be much larger than the placebo units
- But the pre-treatment fits may be different for different control units
- We may just throw out those control units, or
- Abadie et al. (2010) suggests to look a the distribution of ratio of post vs pre-treatment fit
- Yes, we look at the whole distribution, not only p-values.
- If you insist on p-values, you just count how many control units had worse ratio than the treated unit

Placebos

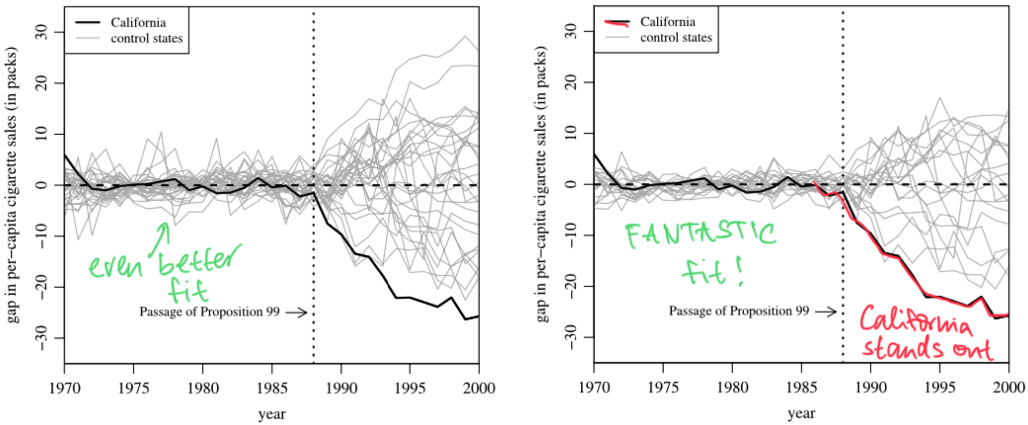
Not a sampling based inference!





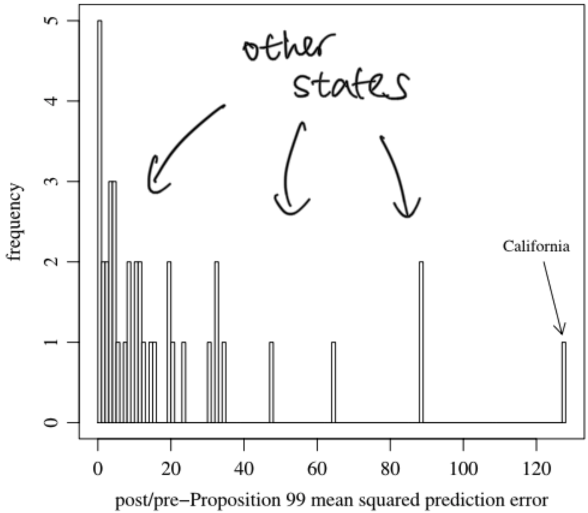
Abadie, Diamond and Hainmueller (2010)

Placebos



Abadie, Diamond and Hainmueller (2010)

Inference



Abadie, Diamond and Hainmueller (2010)

- If the fit is poor in the pre-intervention period. Do not do SCM, do something else.
- Small T_0 and large $J \rightarrow$ risk of overfitting
- Homogenise your pool of potential controls. Make them similar to the control unit.
- Again make comparison more plausible.

But why not regression instead?

Predictors X_0 (with intercept) are used to predict $y_{0,t}$ (post intervention outcomes for J control units at time $t \in T_0 + 1, \ldots, T$):

$$\hat{\beta}_{OLS,t} = (X_0^T X_0)^{-1} X_0^T y_{0,t}$$

$$\underbrace{X_1}_{1\times K} \underbrace{\hat{\beta}_{OLS,t}}_{K\times 1} = \underbrace{X_1 (X_0^T X_0)^{-1} X_0^T}_{w^T \equiv \text{OLS weights}} y_{0,t} = \underbrace{w^T}_{1\times J} \underbrace{y_{0,t}}_{J\times 1}$$

Let us denote $Y_0 = \begin{bmatrix} y_{0,T_0+1} & y_{0,T_0+2} & \cdots & y_{0,T} \end{bmatrix}$ which is $J \times (T - T_0)$ matrix.

$$\underbrace{\hat{B}_{OLS}}_{K\times(T-T_0)} = (\underbrace{X_0^T}_{K\times J} \underbrace{X_0}_{J\times K})^{-1} \underbrace{X_0^T}_{K\times J} \underbrace{Y_0}_{J\times(T-T_0)}$$

$$\underbrace{X_1}_{1\times K} \underbrace{\hat{B}_{OLS}}_{K\times(T-T_0)} = \underbrace{X_1 (X_0^T X_0)^{-1} X_0^T}_{w^T \equiv \text{OLS weights}} Y_0 = \underbrace{w^T}_{1\times J} \underbrace{Y_0}_{J\times(T-T_0)}$$

But why not regression instead?

TABLE 2
Synthetic Control Weights for West Germany

Australia	—
Austria	0.42
Belgium	—
Denmark	—
France	—
Greece	—
Italy	—
Japan	0.16
Netherlands	0.09
New Zealand	—
Norway	—
Portugal	—
Spain	—
Switzerland	0.11
United Kingdom	—
United States	0.22

TABLE 3
Regression Weights for West Germany

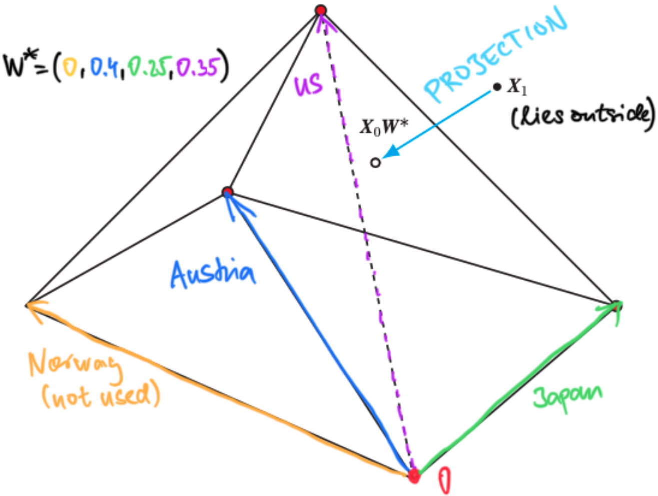
Australia	0.12
Austria	0.26
Belgium	0.00
Denmark	0.08
France	0.04
Greece	−0.09
Italy	−0.05
Japan	0.19
Netherlands	0.14
New Zealand	0.12
Norway	0.04
Portugal	−0.08
Spain	−0.01
Switzerland	0.05
United Kingdom	0.06
United States	0.13

weights negative (?)

Abadie (2021)

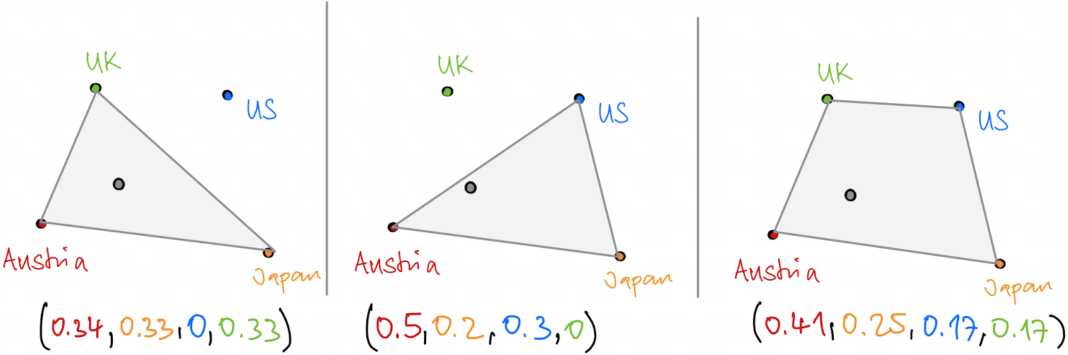
- From OLS we have also weights (!)
- May be negative $\rightarrow$ difficult to interpret
- OLS weights are not sparse
- Sparsity is nice for interpretation

Sparsity?



(modified) Abadie (2021)

Non-uniqueness



Induce sparsity (penalized estimator)

We may induce the sparsity, so penalize for large differences.

$$\arg\min_{w} \underbrace{\left(\sum_{h=1}^{k} v_h \cdot \left(X_{h1} - \sum_{j=2}^{J+1} w_h \cdot X_{hj}\right)^2\right)^{\frac{1}{2}}}_{\text{Regular SCM}} + \lambda \underbrace{\left(\sum_{j=2}^{J+1} w_h \sum_{h=1}^{k} v_h \cdot (X_{h1} - X_{hj})^2\right)^{\frac{1}{2}}}_{\text{Penalty for non-sparse solution}}$$

We are in between the two extreme cases:

- $\lambda \to 0$ - synthetic control method
- $\lambda \to \infty$ - nearest neighbor matching

Compare it to LASSO – in SCM we do care about the maginitude of the estimated weights - they carry important information.

Choice of variables

- No post-treatment outcomes!
- Use outcomes only?? It is easy to interpret, but some covariates may be important too.
- Out-of-sample prediction trick (like for the weights!)
- The general model building guidance applies: simplicity vs fit. We wish to have parsimonious model. One that fits well but does not overfit.

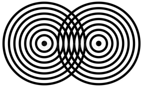
Magnitude of the effect

- How big of an effect would you expect?
- It also depends on how precisely the outcome is measured
- The bigger the noise the more difficult would be to extract the signal from the noise
- Effect needs to be **substantial** so that we can capture it from noisy data
- In case that no appropriate donors exists, you may be interested in modelling differences of growth of outcome instead.
- Or not.

No anticipation

- what if the policy intervention is anticipated
- forward looking agents react in advance
- this would induce a bias
- what could we possibly do about that(?)
- move the time period of intervention back in time
- e.g.it may be the announcement of the policy that matters, not the implementation

No interference



- spillover effects? we may wish to remove some control units
- but this clases with having good donor pool of units
- contextual knowledge about spillovers may inform us about the sign of a potential bias

Advantages

- No extrapolation is made
- The weights make it transparent and easy to interpret
- We know exactly how much each control unit contributes
- Weights are non-negative (unlike for OLS)
- You can fix the weights before the change has occurred.
- Thus you avoid specification fishing.
- You don't need many units, but the right units
- You are relatively close to the data $\rightarrow$ the method is simple

We keep getting back to the most important question:

What do we need to do in order to have a meaningful comparison?

What do many of these methods (RDD, DiD, SCM) have in common??

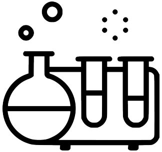
[dramatic pause]

They are very visual.

Professional graphics sells. Make sure to produce beautiful graphs. (See the works of Jonathan Schwabish on how to make great visualizations).

- Schwabish, Jonathan A. "An economist's guide to visualizing data." Journal of Economic Perspectives 28.1 (2014): 209-34.
- Schwabish, Jonathan. Better presentations. Columbia University Press, 2016.
- Schwabish, Jonathan. Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks. Columbia University Press, 2021.

Synthetic controls and experimentation



- What is the impact of a new policy?
- We can only experiment on larger units (say cities).
- We choose some units (cities) and weight them to construct synthetic treatment unit, that resembles the population of interest.
- Construct synthetic control unit for this synthetic treatment unit
- And compare them. Yes, that's it.
- This has been used in the industry for a longer time.
- Abadie and Zhao (2021) worked out the math.

Implementation

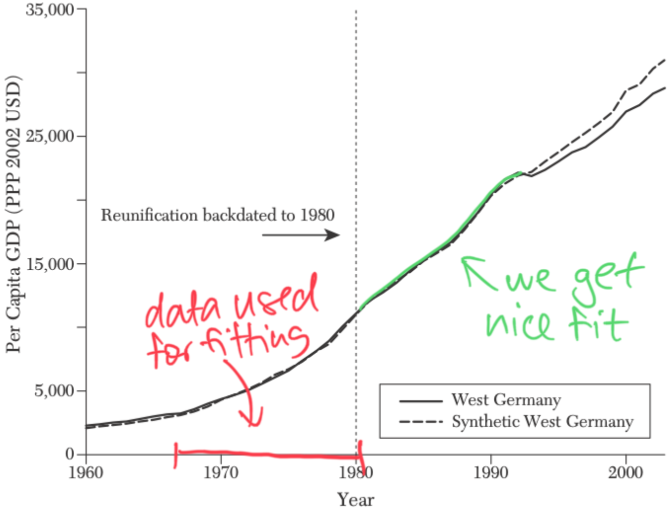
- R - `gsynth` and `tidysynth`
- STATA - `synth` package

Diagnostics and Robustness analysis

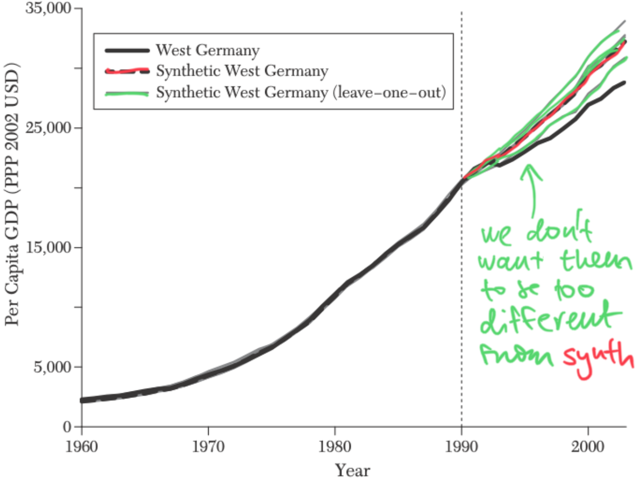
How do we know a model is fine.

- Important part
- backdating - try to implement the treatment date at an earlier stage
- leave-on-out reanalysis

Backdating


35,000
25,000
15,000
5,000
0
Per Capita GDP (PPP 2002 USD)
1960
1970
1980
1990
2000
Year
Reunification backdated to 1980
West Germany
Synthetic West Germany
data used for fitting
we get nice fit

Leave-one-out reanalysis



Abadie, Diamond and Hainmueller (2010)

Assuming a linear factor model: If you manage to match controls and outcome in the pre-treatment periods ($T = 1, ..., T_0$) then you can bound the bias (Abadie, Diamond, and Hainmueller 2010).

$$Y_{jt}^N = \delta_t + \theta_t Z_j + \lambda_t \mu_t + \varepsilon_{jt}$$

$$\mathrm{E}\left[Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N \right]$$

- Y_{jt}^N - counterfactual outcome under non-treatment for unit j in time t (unobserved for $j = 1$ and $t > T_0$)
- δ_t - time trend
- Z_j - vector of observed predictors
- μ_j - vector of unobserved predictors
- θ_t, λ_t - coefficients
- ε_{jt} - zero mean individual transitory shocks

The bias $E\left[Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N\right]$ bound under this linear factor model

$$Y_{jt}^N = \delta_t + \theta_t Z_j + \lambda_t \mu_t + \varepsilon_{jt}$$

- is based on the fact that pre intervention fit is perfect $X_1 = X_0 W^*$
- decreases with larger values of T_0
- increases with the number of units in the donor pool J
- increases if unobserved μ_j greatly differ from μ_1
- increases with the dimension of μ_j
- is based on a linear model, so if the true model is non-linear then the bias formula does not hold

lesson to take:

the comparison units should be chosen carefuly.

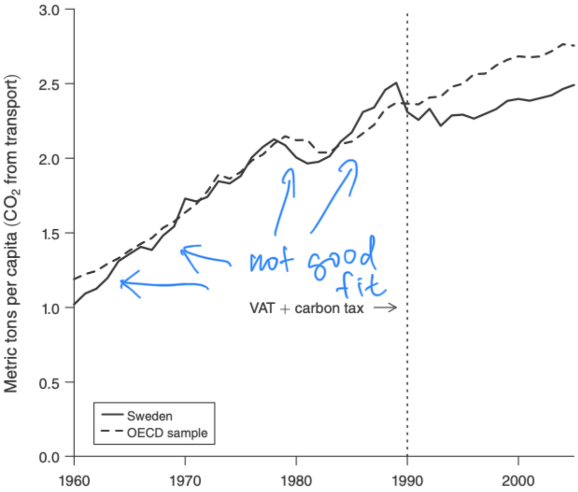
More examples

- carbon tax in Sweden (Andersson 2019)
- universal cash transfers and labor markets (Jones and Marinescu, 2022)
- right-to-carry laws (Donohue, Aneja, and Weber 2019)
- legalized prostitution (Cunningham and Shah 2018)
- immigration policy (Bohn, Lofstrom, and Raphael 2014),
- corporate political connections (Acemoglu et al. 2016),
- taxation (Kleven, Landais, and Saez 2013),
- organized crime (Pinotti 2015),
- effects of immigration (Borjas 2017)
- minimum wages (Allegretto et al. 2017, Jardim et al. 2017)

Even more examples

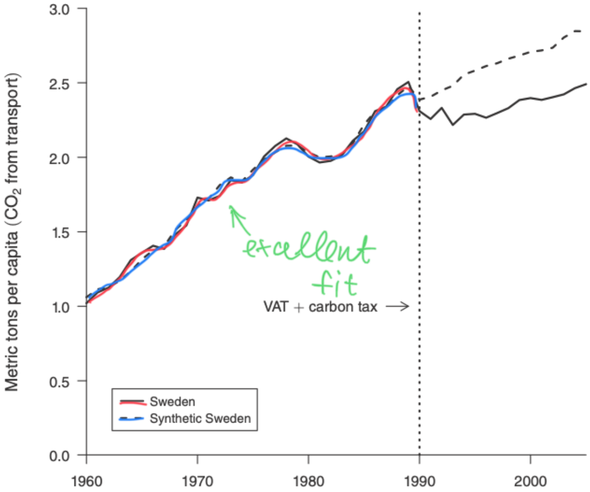
- social sciences, biomedical disciplines, engineering, etc. (see, e.g., Heersink, Peterson, and Jenkins 2017; Pieters et al. 2017)
- Bill & Melinda Gates Foundation's Intensive Partnerships for Effective Teaching program (Gutierrez, Weinberger, and Engberg 2016).

Example: Carbon tax in Sweden



Andrersson (2019)

Synthetic Sweden



Andrersson (2019)

Balance

Variables	Sweden	Synth. Sweden	OECD sample
GDP per capita	20,121.5	20,121.2	21,277.8
Motor vehicles (per 1,000 people)	405.6	406.2	517.5
Gasoline consumption per capita	456.2	406.8	678.9
Urban population	83.1	83.1	74.1
CO_2 from transport per capita 1989	2.5	2.5	3.5
CO_2 from transport per capita 1980	2.0	2.0	3.2
CO_2 from transport per capita 1970	1.7	1.7	2.8

Andrersson (2019)

Synthetic weights

Country	Weight	Country	Weight
Australia	0.001	Japan	0
Belgium	0.195	New Zealand	0.177
Canada	0	Poland	0.001
Denmark	0.384	Portugal	0
France	0	Spain	0
Greece	0.090	Switzerland	0.061
Iceland	0.001	United States	0.088

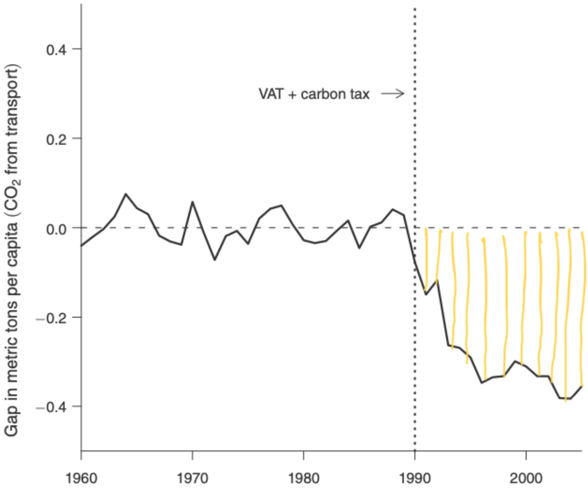
$w \geq 0$

sparse

$\sum_i w_i = 1$

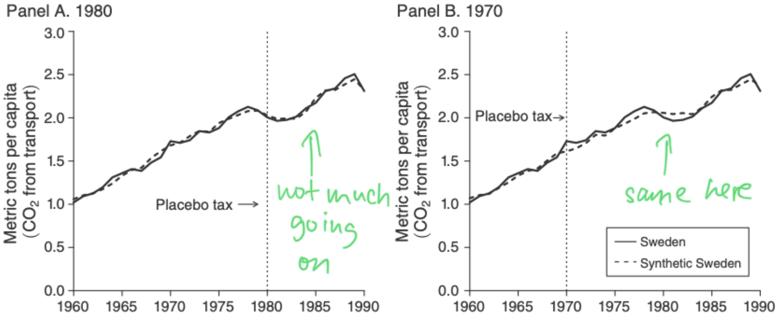
Andrersson (2019)

Effects



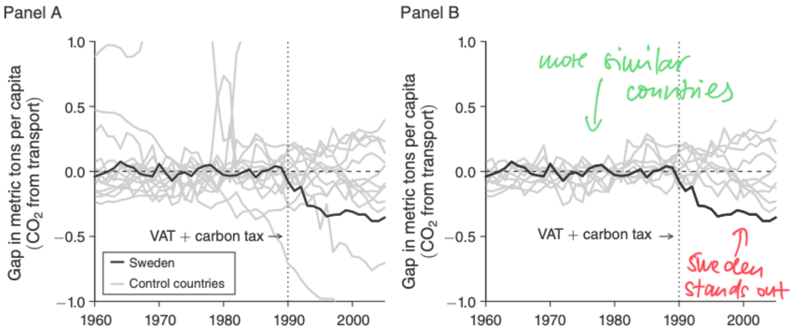
Andrersson (2019)

Placebos



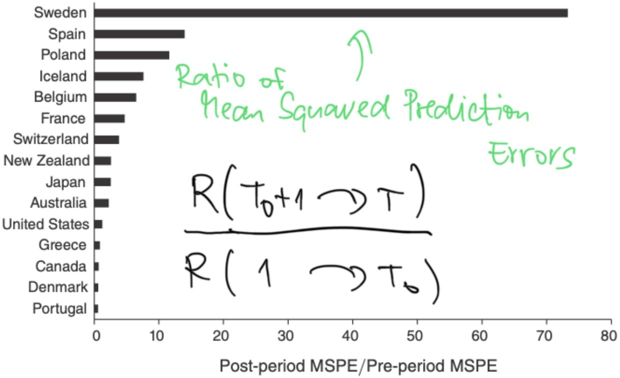
Andrersson (2019)

Statistical Inference



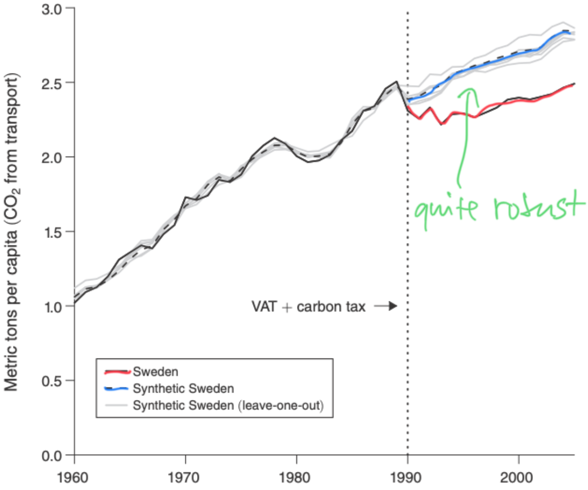
Andrersson (2019)

Ratio of post/pre intervention fit



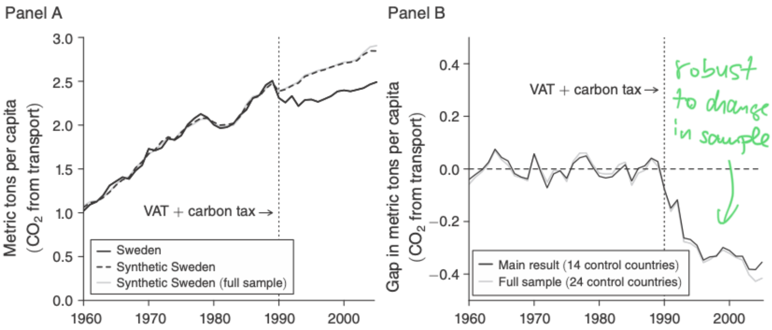
Andrersson (2019)

Leave-one-out analysis



Andrersson (2019)

Robustness to change in sample



Andrersson (2019)

How often does your favorite econometric technique get featured in The Washington Post or The Wall Street Journal?

The Washington Post
Democracy Dies in Darkness

This article was published more than **7 years ago**

ECONOMIC POLICY

Seriously, here's one amazing math trick to learn what can't be known

By Jeff Guo
October 30, 2015

THE WALL STREET JOURNAL.

SUBSCRIBE | SIGN IN

ECONOMICS BLOG

How an Analysis of Basque Terrorism Helps Economists Understand Brexit

A method pioneered by an MIT professor has also been used to estimate the economic effect of a tobacco ban, German reunification, legalization of prostitution and gun rights

By Jason Douglas Follow
Nov. 7, 2018 5:37 am ET

SHARE TEXT

A method developed more than a decade ago to assess the cost of political violence in the Basque country has become a key tool for economists trying to figure out the cost of Brexit.

Guo, Jeff. 2015. "Seriously, Here's One Amazing Math Trick to Learn What Can't Be Known." Washington Post, October 30.

Douglas, Jason. 2018. "How an Analysis of Basque Terrorism Helps Economists Understand Brexit." WallStreet Journal, November 7.

- SCM is new
- It is very popular and constantly getting more traction
- Much more will be done in the next few years
- It became a standard in econometrics toolbox

Thank you for your attention!

Main References

- Original paper: Abadie, Alberto, and Javier Gardeazabal. "The economic costs of conflict: A case study of the Basque Country." American economic review 93.1 (2003): 113-132.
- Paper where theory is worked out: Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." Journal of the American statistical Association 105.490 (2010): 493-505.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Comparative politics and the synthetic control method." American Journal of Political Science 59.2 (2015): 495-510.
- Recent review article: Abadie, Alberto. "Using synthetic controls: Feasibility, data requirements, and methodological aspects." Journal of Economic Literature 59.2 (2021): 391-425.
- Instructive video by inventor of SCM himself `https://www.youtube.com/watch?v=nKzNp-qpE-I`
- Similar, slightly longer video also by Abadie at 2021 NBER Summer Institute lecture series `https://www.youtube.com/watch?v=T2p9Wg650bY`
- Abadie, Alberto, and Jinglong Zhao. "Synthetic controls for experimental design." arXiv preprint arXiv:2108.02196 (2021).
- Chapter 10 in S.Cunningham's book: `https://mixtape.scunning.com/synthetic-control.html`
- It is not often that WSJ writes about econometric methods: `https://www.wsj.com/articles/how-an-analysis-of-basque-terrorism-helps-economists-understand-brexit-1541587068`
- Andersson, Julius J. "Carbon taxes and CO 2 emissions: Sweden as a case study." American Economic Journal: Economic Policy 11.4 (2019): 1-30.