

## 6. Synthetic control methods

LPO 8852: Regression II

Sean P. Corcoran

### Synthetic control method - introduction

Synthetic control methods are often used when there is only *one* treated unit observed before and after treatment and no clear comparison unit.

- A context where one might like to use difference-in-differences but potential comparison units are quite different from the treated unit, such that the parallel trends assumption fails to hold.
- The treated unit is often at a high level of aggregation (e.g., country, region, state) but could be a smaller unit (e.g., school or firm).
- Abadie and Gardeazabal (2003) is the classic reference—on the impact of terrorism on economic activity.
- Abadie et al. (2010) elaborate on the methods in the context of an anti-smoking law in California.
- Abadie (2021) is an excellent survey—highly recommended.

# Motivating example

A classic paper by Card (1990) looks at the effect of the Mariel Boatlift from Cuba in 1980 on the Miami labor market. Cuban immigrants increased the size of the local labor force by 7%. What affect did this have on the wages of less-skilled native workers?

- Used Current Population Survey data on unemployment of native born workers.
- Estimated a standard difference-in-differences using similar workers in Atlanta, Los Angeles, Houston, and Tampa.
- The selection of comparisons was arguably *ad hoc*.
- This study famously found no impact of immigration on local unemployment.

## Synthetic control

**Synthetic control methods** optimally choose a set of weights that—when applied to a group of corresponding untreated units—provides a counterfactual path for the unit that received the treatment.

- This weighted group is the “synthetic unit” and stands in for what would have happened to the aggregate treated unit had the treatment not occurred.

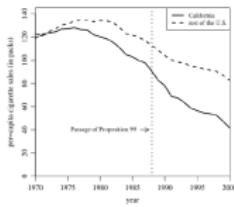


Figure 1. Trends in per capita cigarette sales: California vs. the rest of the United States.



Figure 2. Trends in per capita cigarette sales: California vs. synthetic California.

Figures 1 and 2 from Abadie et. al. (2010)

# Synthetic control

Advantages:

- Doesn't require large sample of treated and untreated cases.
- Selection of comparison units—and their exact weighting—is explicit.
- Doesn't extrapolate as is done in regression; estimates are always a weighted average of other non-treated units.
- Only need pre-treatment data to construct weights.

Disadvantage:

- Requires a sufficient number of pre-treatment observations to identify a "good" synthetic control.

## Formalization

- $Y_{jt}$  is the outcome of interest for unit  $j$  in period  $t$
- There are  $J + 1$  units and unit 1 is treated. The other  $J$  units are the **donor pool**.
- There are  $T$  periods,  $1, \dots, T_0$  before treatment and  $T_0 + 1, \dots, T$  after
- We may also observe predictors of  $Y_{jt}$  that are time-varying:  $X_{jkt}$  ( $k = 1, \dots, K$ )

## Potential outcomes

- Can think about potential outcomes  $Y_{jt}(1)$  and  $Y_{jt}(0)$  so that the treatment effect for unit  $j$  is  $\tau_{jt} = Y_{jt}(1) - Y_{jt}(0)$ . Note  $\tau$  is *time varying*.
- As always, the challenge is to estimate  $Y_{jt}(0)$  for the treated case in the treated period(s). That is, how would  $Y_{jt}$  have evolved in the absence of treatment?
- **Synthetic control:** finding a weighted combination of units in the donor pool to approximate  $Y_{jt}(0)$ :

$$\hat{Y}_{jt}(0) = \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

where  $w_j^*$  is a set of optimally chosen weights.

## Potential outcomes

- So then the estimated treatment effect for unit 1 in time  $t$  is:

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

- The goal is to identify a weighted combination of units in the donor pool that approximates  $Y_{1t}(0)$  in the *pre-treatment* period. The assumption is that if this group tracked the treated observation in the pre-period, it would continue to do so in the post-period, in the absence of treatment.

## Finding weights

So how do we obtain these weights? What criteria do we use?

- Usually, weights are constrained to be non-negative and sum to one.
- One possibility would be to just use equal weights  $w_j = 1/J$ :

$$\hat{\tau}_{1t} = Y_{1t}(1) - \frac{1}{J} \sum_{j=2}^{J+1} Y_{jt}$$

- Or a population-weighted version where  $w_j^{pop}$  is the size of unit  $j$  as a fraction of the total donor pool:

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

These often do not perform that well.

## Finding weights

Abadie and Gardezabal (2003) and Abadie et al. (2010) propose choosing weights so that the synthetic control best resembles the *pre-treatment values* of the treated unit for predictors  $X$  (all variables not affected by the treatment). I.e. minimize the distance between  $X$  for unit 1 and the weighted  $X$  for untreated units.

$$||\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}|| = \left( \sum_{h=1}^k v_h (X_{h1} - w_2 X_{h2} - \dots - w_{J+1} X_{hJ+1})^2 \right)^{1/2}$$

This involves first finding constants  $v_k$  that reflect the relative importance of the predictors  $X_{k1}$  as a predictor of  $Y_{1t}(0)$ .

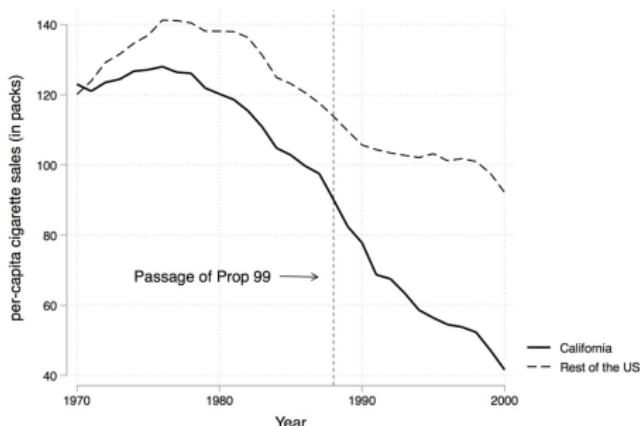
## Finding weights

In practice the predictors  $X$  above include pre-treatment outcomes ( $Y$ ), pre-treatment predictors ( $X$ ), or both. Often these are averages over the pre-treatment periods.

According to Abadie (2021), *sparsity* is typical of the synthetic control method—i.e., weights tend to be concentrated on a small number of units.

### Example: Abadie et al. (2010)

The impact of Proposition 99 on per-capita cigarette sales. Prop 99 increased cigarette taxes and funded other anti-smoking initiatives.



## Example: Abadie et al. (2010)

Per-capita cigarette sales for California and “synthetic California”

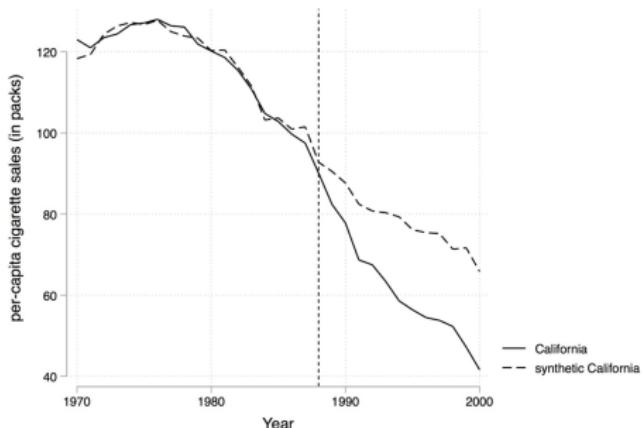


Figure 10.2: California cigarette sales vs synthetic California

## Example: Abadie et al. (2010)

Variables used to find optimal weights—includes both predictors of cigarette sales *and* pre-treatment values of cigarette sales:

Table 10.1: Balance table

Variables	Real California	Synthetic Calif.	Avg. 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

All variables except lagged cigarette sales are averaged for the 1980–1988 period. Beer consumption is averaged 1984–1988.

## Example: Abadie et al. (2010)

Another nice way to show pre- and post-treatment gap in the outcome:

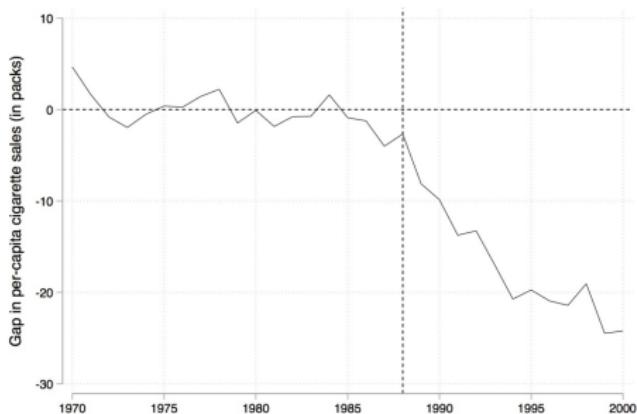


Figure 10.3: Gap in cigarette sales for estimation pre and post treatment

## Inference

The graph is compelling, but how can we say whether the post-treatment difference is *significant* or not?

- Inference is atypical here since we only have one observation (per year) on the treated and synthetic control group.
- Similarly, cannot formally test for baseline differences in Table 10.1 above since there is only one observation in each group.
- Abadie et al. (2010) propose *placebo-based* inference:
  - ▶ It's possible that the SCM can yield differences even when there is no treatment effect.
  - ▶ Apply the SCM to *every* unit in the donor pool, treating that unit as the "treated" case, and calculate the (placebo) treatment effect.
  - ▶ How *unusual* is the (actual) treatment effect relative to these?

## Example: Abadie et al. (2010)

All placebo cases overlaid with the actual treatment case (in bold):

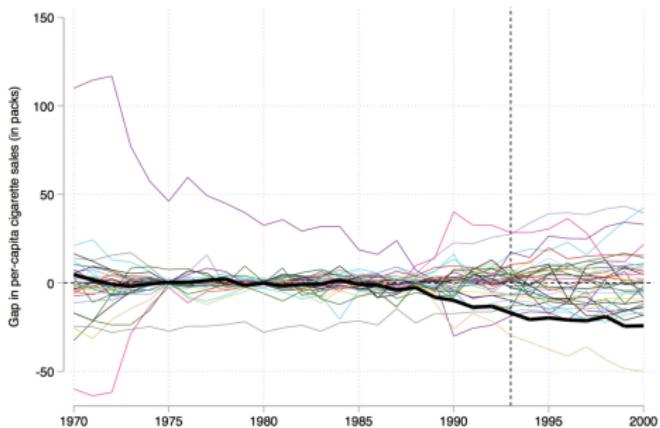
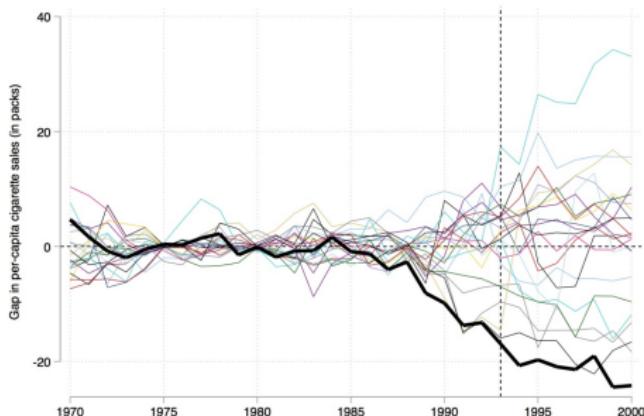


Figure 10.4: Placebo distribution using all units as donor pool

## Example: Abadie et al. (2010)

Dropping units with very poor fit pre-treatment can help this graph:



# Inference

In addition to visual evidence, we can calculate a test statistic:

- ① Apply SCM to each unit in the donor pool
- ② Calculate the root mean square prediction error (RMSPE) in the pre-treatment period (i.e., how well does the donor pool track the “treated” case in the pre-period?)
- ③ Calculate the RMSPE in the post-treatment period (i.e., how closely does the donor pool track the “treated” case in the post-period? If there is an effect, these should diverge).
- ④ Compute the *ratio* of post- to pre-RMSPE.
- ⑤ Sort this ratio from highest to lowest.
- ⑥ Calculate the rank of the (actual) treatment unit ratio in this distribution and calculate a “*p-value*” (Rank/Total).

## Applications: education and health

- Shores, Candelaria & Kabourek (2023): effects of school finance reforms. Reform states contrasted with a weighted average of non-reform states.
- Gutierrez, Weinberger, & Engberg (2016): Gates Foundation Intensive Partnerships for Effective Teaching program.
- Bifulco, Rubenstein, & Sohn (2017): “Say Yes to Education” promise scholarship in Syracuse.
- Dave et al. (2021): impact of Trump rally in Tulsa on COVID-19 spread.
- Trejo et al. (2021): academic and psychosocial effects of Flint water crises (e.g., anxiety and depression)

"Say Yes to Education" was a promise scholarship program implemented in Syracuse NY in 2008. It offers full tuition scholarships for public and private universities, coupled with wraparound support services in schools.

- They consider two donor pools, a restricted one (judged similar to Syracuse in important ways) and a comprehensive one (all urban or suburban districts in NYS). See Table 2.
- They try six alternative combinations of pre-treatment year outcome means to find weights. See Table 1. They also use average shares FRPL, African-American, and Hispanic (over the whole pre-period).
- There are two different outcomes: enrollment and graduation rates. Weights can and will differ depending on the outcome.

# Bifulco et al. (2017)

Alternative combinations of pre-treatment years to find weights (Table 1):

**Table I.** Alternative Specifications of Pretreatment Years.

Specification	Description
1	First and last year of pretreatment periods
2	First, middle, and last year of pretreatment periods
3	Middle and last year of pretreatment periods
4	Last pretreatment year and the average of outcomes in all other pretreatment years
5	Each pretreatment year
6	Each year from the middle to the end of the pretreatment periods

**Table 2.** Assignment of Weights (Enrollment Analysis).

District Name	Assigned Weights					
	Specif. 1	Specif. 2	Specif. 3	Specif. 4	Specif. 5	Specif. 6
<b>Panel A: Restricted donor pool</b>						
Albany CSD	.000	.021	.000	.000	.005	.000
Brentwood UFSD	.000	.000	.000	.000	.116	.129
Buffalo CSD	.000	.078	.022	.034	.029	.000
Niagara Falls CSD	.484	.288	.498	.499	.411	.404
Rochester CSD	.502	.406	.479	.467	.438	.467
Utica CSD	.014	.207	.001	.000	.000	.000
<b>Panel B: Comprehensive donor pool</b>						
Albany CSD	.000	.000	.000	.000	.005	.000
Brentwood UFSD	.000	.000	.000	.000	.116	.134
Buffalo CSD	.117	.174	.091	.065	.029	.004
Elmira CSD	.000	.000	.307	.148	.000	.000
Hopevale UFSD	.101	.197	.053	.021	.000	.000
Mount Vernon CSD	.000	.061	.000	.000	.000	.000
Niagara Falls CSD	.248	.000	.124	.324	.411	.401
Rochester CSD	.386	.341	.425	.442	.438	.461
Smythtown CSD	.149	.156	.000	.000	.000	.000
Utica CSD	.000	.069	.000	.000	.000	.000

Note. Specif., CSD, and UFSD denote "specification," "city school district," and "union free school district," respectively. Districts that do not appear in the table do not receive positive weights equal to or greater than 0.001 in any of the specifications. Restricted donor pool includes 22 districts, whereas comprehensive donor pool includes 275 districts.

## Bifulco et al. (2017)

### Results: enrollment outcome (Figure 2)

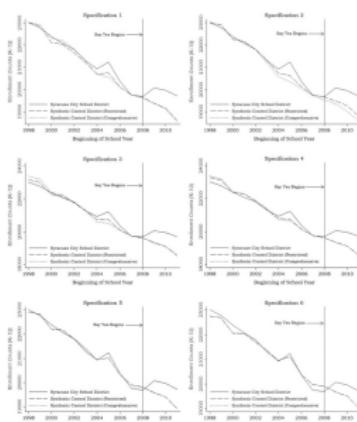


Figure 2. Trends in enrollment by model specifications. See Table 1 for description of pretreatment years included in each specification.

## On inference:

- Following Abadie et al. (2010), they use a permutation test to assess the likelihood that the effect found would occur by chance.
- They apply SCM for each unit in the donor pool and compare the “real” effect to these “placebo” effects.
- For each iteration of SCM, they compute RMPSE: the square root of the average of the squared prediction errors across years in the pre- and post-treatment periods. The ratio of these is a test statistic.
- They calculate the percentage of test statistics that are as large as that obtained for Syracuse. Interpreted as the probability of obtaining effect estimates this large if the true treatment effect were zero.

## Bifulco et al. (2017)

## Results: enrollment effects

**Table 3.** Estimated Effects on K-12 Enrollments, RMSPE, and *p*-value.

Specification	Year 1	Year 2	Year 3	Year 4	RMSPE	<i>p</i> -value
<b>Panel A: Restricted donor pool</b>						
Specification 1	30	704	789	1,110	214.16	.087
Specification 2	24	576	676	889	220.09	.217
Specification 3	69	739	840	1,166	209.04	.044
Specification 4	64	730	839	1,165	227.19	.130
Specification 5	-114	500	560	795	114.27	.304
Specification 6	-216	405	445	672	161.23	.130
<b>Panel B: Comprehensive donor pool</b>						
Specification 1	87	693	859	1,132	252.35	.076
Specification 2	147	715	920	1,117	318.44	.243
Specification 3	76	702	845	1,075	323.72	.098
Specification 4	67	713	840	1,121	280.36	.091
Specification 5	-114	500	560	795	114.27	.562
Specification 6	-213	404	445	668	156.37	.120

Note: Restricted donor pool includes 22 districts, whereas comprehensive donor pool includes 275 districts. Years 1–4 correspond to the effect estimates. *p*-value implies a probability of getting a post/pretreatment RMSPE ratio as large as the post/pretreatment RMSPE ratio of Syracuse if one assigns the treatment at random in the data. Specifications are the same as in Table 1. Pretreatment period includes years 1998–2007. All models are run with percent Black, percent Hispanic, and percent free lunch eligible as covariates. RMSPE = root mean squared prediction error.

# Synthetic control in Stata

Install the synth package (or synth\_runner for added functionality). Note the package assumes the cross-sectional and time variables have been set using tsset *unit time*.

`synth depvar predictorvars, options`

Gives you:

- Weights
- RMSPE
- Balance table (for *predictorvars*)
- Figure

synth2 is a more recent wrapper for the synth package. See Yan and Chen (2023).

## Synthetic control in Stata

Options:

- `trunit()`: ID of treatment unit
- `trperiod()`: time period when treatment occurs
- `mspeperiod()`: pre-treatment time period used to identify synthetic control (MSPE = mean squared prediction error)
- `resultsperiod()`: time period for which results are reported, displayed in a figure, saved (can be pre- and post-treatment)
- `unitnames()`: variable containing names of cross-sectional unit
- `figure`: request figure displaying the results
- `keep(file)`: designate a filename for saving results
- Others (see help menu)

## Example: from the *Mixtape*

What was the impact of prison construction in Texas on the incarcerated population? In 1993, Gov. Richards funded a significant expansion of prisons in Texas in response to an earlier court ruling about prison over-crowding. Did these new prisons increase the incarcerated population, perhaps due to a curtailment of parole? Did expanded prison capacity have a disproportionate impact on Black males?

- See annotated do file on Github
- Use other states as a synthetic control, estimate treatment effect
- Manually do placebo inference by running synth for all states

The above .do file also shows the use of synth\_runner and synth2.