

7. 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 want 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 single school or firm (for example).
- 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 (see reading list).

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 local labor force by 7%, and potentially reduced the wages of less-skilled native workers.

- Used Current Population Survey data on unemployment of native born workers.
- A standard difference-in-differences using workers in Atlanta, Los Angeles, Houston, and Tampa.
- Selection of comparisons was arguably *ad hoc*, and standard errors reflect sampling variance only, not uncertainty about the comparison group.
- This study found no impact of immigration on local unemployment.

Synthetic control

Synthetic control methods optimally choose a set of weights, which—when applied to a group of corresponding untreated units—produces an optimally estimated counterfactual to 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.

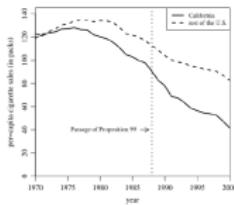


Figure 1. Trends in per capita cigarette sales: California vs. the rest of the United States.



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.

Formalization

- Y_{jt} is the outcome of interest for unit j in period t
- There are $J + 1$ units and (without loss of generality) 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 $\tau_{jt} = Y_{jt}^1 - Y_{jt}^0$. Notice it 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^0 in the *pre-treatment* period. The assumption is that if this group tracks the treated observation in the pre-period, it would continue to do so in the post-period.

Finding weights

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

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

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

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

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

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.

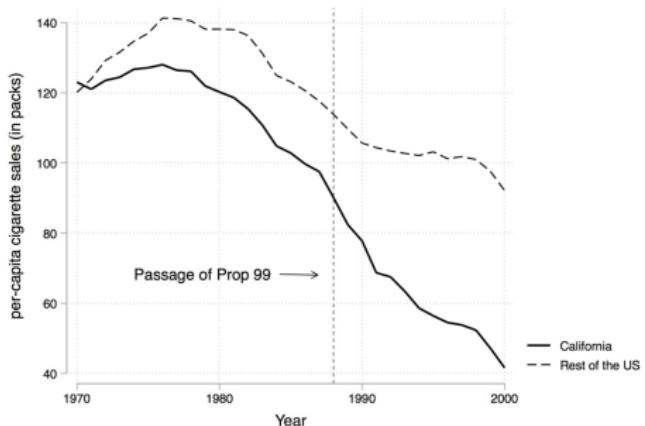
- This involves constants v_k that reflect the relative importance of the predictors X_{k1} as a predictor of Y_{1t}^0 .

In practice can use pre-treatment outcomes, 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”

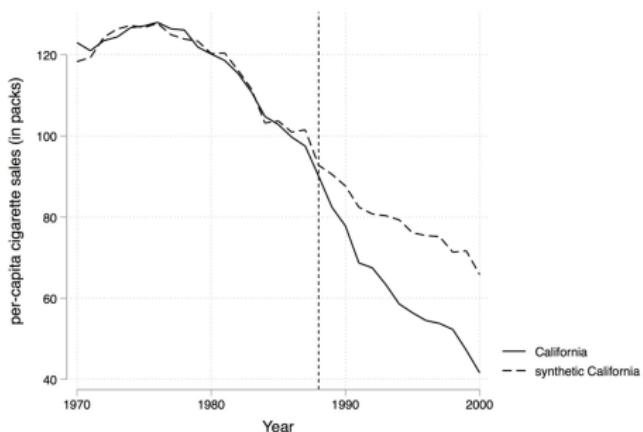


Figure 10.2: California cigarette sales vs synthetic California

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

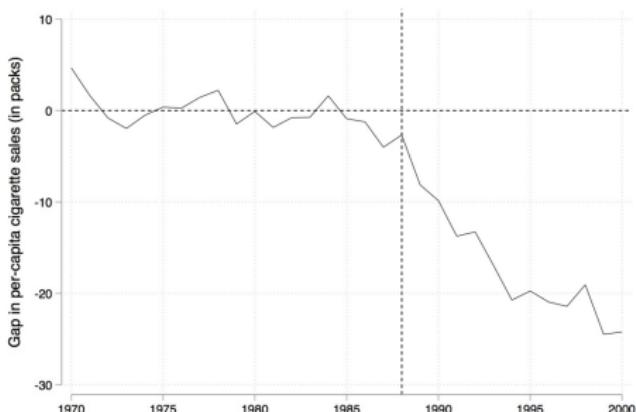


Figure 10.3: Gap in cigarette sales for estimation pre and post treatment

Inference

The graph is compelling, but how can we say whether the 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 method yields 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:

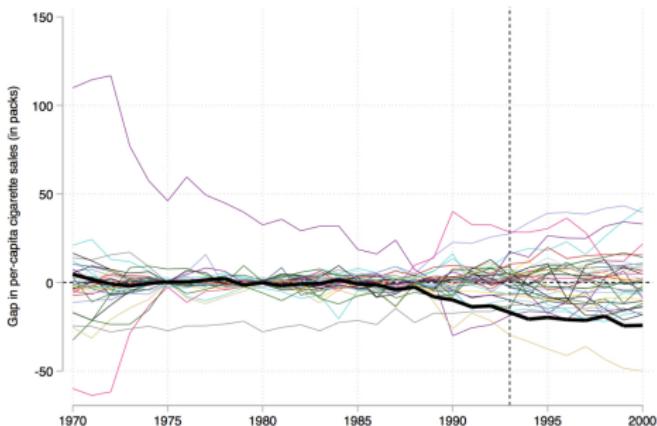


Figure 10.4: Placebo distribution using all units as donor pool

Example: Abadie et al. (2010)

Dropping units with very poor fit pre-treatment.

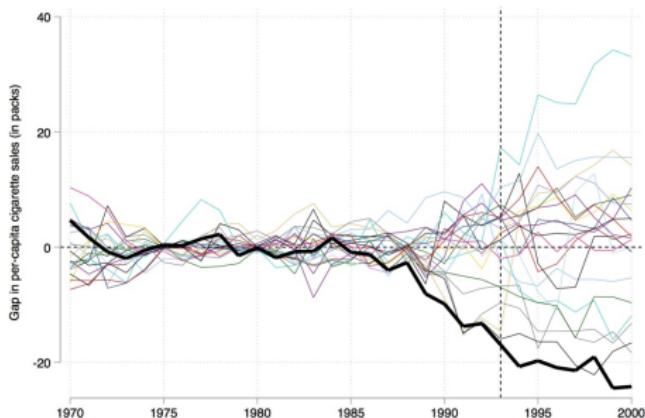


Figure 10.5: Pre-Proposition 99 RMSPE ≥ 2 times Pre-Pop 99 RMSPE for CA

Inference

In addition to visual evidence can calculate a test statistic:

- ① Apply SCM to each unit in the donor pool
- ② Calculate the root mean square prediction error in the pre-treatment period (i.e., how well does the donor pool track the “treated” case in the pre-period?)
- ③ Calculate the root mean square prediction error 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-RMSE.
- ⑤ Sort this ratio from highest to lowest.
- ⑥ Calculate the rank of the (actual) treatment unit ratio distribution, and calculate a “p-value” (Rank/Total).

Applications: education

- Candelaria & Shores (forthcoming): effects of school finance reforms. Reform states contrasted with a weighted average of non-reform states. Uses "ridge-augmented" SCM that adjusts for any remaining pre-treatment differences.
- Gutierrez, Weinberger, & Engberg (2016): Gates Foundation Intensive Partnerships for Effective Teaching program.
- Bifulco, Rubenstein, & Sohn (2017): "Say Yes to Education" promise scholarship in Syracuse.
- Ridley & Terrier (2018): impact of lifting charter school cap in Massachusetts. Combined with IV and DD.

Bifulco et al. (2017)

Treated school district: Syracuse CSD

- 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).
- They try alternative specifications of pre-treatment years when finding weights.
- There are two different outcomes (enrollment and graduation rates); weights will differ depending on the outcome.

On variables used to find weights:

Researchers must also choose the linear combinations of pretreatment outcome observations the synthetic control should seek to match. One natural choice is the value of the outcome variable for all the available pretreatment periods. In practice, including all available pretreatment outcomes measures does not necessarily result in a closely matched synthetic control. For instance, some of the pretreatment outcome values might be disproportionately influenced by transitory shocks, and it is not clear that closely matching those values will improve estimates of the counterfactual outcomes for the treated district during the posttreatment period. The existing literature provides little guidance about how to choose the set of linear combinations of pretreatment outcome observation periods.

To explore the sensitivity of effect estimates to the choice of pretreatment outcome observations, we construct synthetic controls using six different sets of pretreatment years, which are detailed in Table 1. To allow for assigning weights to districts that are similar to SCSD in terms of socioeconomic characteristics, we also include in all of our specifications the average shares of free-lunch eligible, African American, and Hispanic students over the entire pretreatment period, which for the enrollment analysis ranges from 1998–1999 to 2007–2008 and for the graduation analysis covers the 1998 through 2004 cohorts. The inclusion of covariates (e.g., the average share of free-lunch eligible) when estimating the weights reduces the likelihood that some districts quite dissimilar demographically to Syracuse could contribute to the counterfactual outcome.

Table 1. 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

Bifulco et al. (2017)

Table 2. Assignment of Weights (Enrollment Analysis).

District Name	Assigned Weights					
	Specif. 1	Specif. 2	Specif. 3	Specif. 4	Specif. 5	Specif. 6
Panel A: Restricted donor pool						
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
Panel B: Comprehensive donor pool						
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.

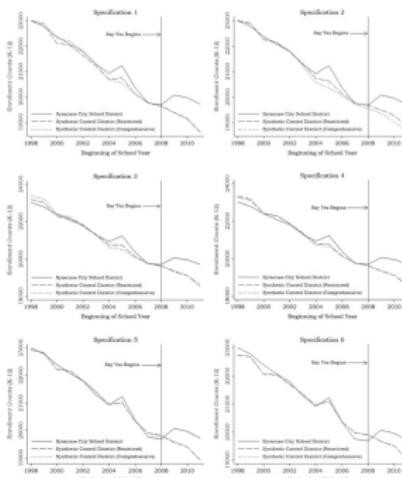


Figure 2. Trends in enrollment by model specifications. See Table 1 for description of pretreatment years included in each specification.

Bifulco et al. (2017)

On inference:

Inference Procedure

Because effect estimates are based on comparison of a single treated unit with a synthetic control, an asymptotic approximation is not well suited for assessing the uncertainty in the estimates. Abadie et al. (2010) propose a permutation test to assess the likelihood that the effect size found in the treated unit is sufficiently unlikely to occur by chance. Specifically, they recommend constructing a synthetic control for each unit in the sample, using the remaining districts as the donor pool for each other district. The true treatment effect in each of these untreated units is presumably zero. By comparing the “effect estimate” in each of the untreated units to that found in the treated unit, one can assess how likely it is to obtain an effect estimate that large by chance.

A stylized example helps to illustrate the procedure. Assume we have 1 treated unit and 99 untreated donor units. The treated unit receives the treatment at time t . We construct a synthetic control for the treated unit and find that outcome value in the treated unit increased by the amount ΔY . We then construct synthetic controls for each of the untreated donor units, assuming that these untreated units also received the treatment at time t , and measure the increase in the outcome values in each, relative to its synthetic control, in the posttreatment period. If we find that none of the 99 untreated district has an increase as large as or larger than ΔY in the posttreatment period, this is analogous to a 99% confidence level the increase in the treated district was not due to chance.

More specifically, for each iteration of the synthetic control procedure, we compute the prediction errors for each year, square them (because the errors can be either positive or negative), and take the square root of the average of the squared prediction errors in the pretreatment period and the square root of the average of the squared prediction errors in the posttreatment period. These are called the pre- and posttreatment root mean squared prediction errors (RMSPEs). The ratio of the posttreatment RMSPE to the pretreatment RMSPE is used as a test statistic. The higher this test statistic, the greater the deviation in the posttreatment period between the district and its synthetic control and, therefore, the stronger the evidence of a policy impact. Calculating the percentage of test statistics from all iterations of this procedure that are as large as or larger than that obtained for Syracuse provides a p -value. Because we assume that the effect of Say Yes in all the donor pool districts is zero, this p -value can be interpreted as the probability of obtaining effect estimates as large as that obtained for Syracuse if the true treatment effect were zero (Abadie, Diamond, & Hainmueller, 2010; Cameron & Miller, 2015).

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
Panel A: Restricted donor pool						
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
Panel B: Comprehensive donor pool						
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.

Bifulco et al. (2017)

Table 5. Assignment of Weights (Cohort Graduation Rate Analysis).

District Name	Assigned Weights				
	Specif. 1	Specif. 2	Specif. 3	Specif. 4	Specif. 5
Panel A: Restricted donor pool					
Albany CSD	.064	.000	.000	.000	.000
Buffalo CSD	.323	.800	.000	.608	.646
East Ramapo CSD	.000	.000	.000	.000	.027
Hempstead UFSD	.000	.021	.000	.000	.134
Hudson CSD	.000	.000	.002	.000	.000
Niagara Falls CSD	.000	.101	.160	.000	.000
Poughkeepsie CSD	.000	.000	.396	.000	.000
Rochester CSD	.351	.000	.000	.157	.000
Schenectady CSD	.261	.078	.442	.232	.219
Panel B: Comprehensive donor pool					
Albany CSD	.042	.000	.000	.008	.012
Buffalo CSD	.679	.477	.368	.683	.000
Binghamton CSD	.000	.000	.000	.000	.131
East Ramapo	.000	.000	.000	.000	.066
Central SD					
Bethpage CSD	.190	.000	.000	.193	.000
Glen Cove CSD	.000	.000	.000	.000	.052
Greenburgh	.000	.000	.000	.000	.019
Greenburgh,					
Hawth.-Cedar Kn. UFSD	.090	.166	.015	.072	.219
Westbury UFSD	.000	.000	.000	.000	.111
Hempstead UFSD	.000	.000	.000	.000	.025
Niagara Falls CSD	.000	.168	.000	.000	.000
Poughkeepsie CSD	.000	.000	.391	.047	.058
Rochester CSD	.000	.189	.226	.000	.269
Watervliet CSD	.000	.000	.000	.000	.002
Westerville CSD	.000	.000	.000	.000	.323
Westbury UFSD	.000	.000	.000	.000	.018

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 236 districts.

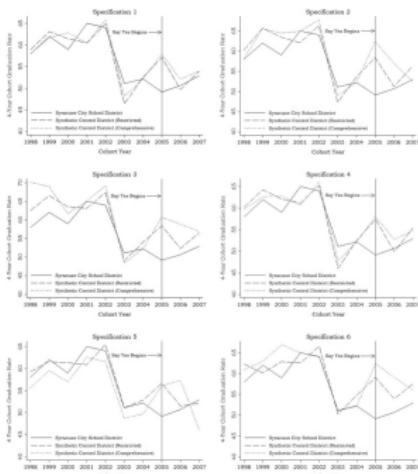


Figure 3. Trends in 4-year cohort graduation rate by model specifications. See Table 1 for description of pretreatment years included in each specification.

Bifulco et al. (2017)

Table 6. Estimated Effects on Cohort Graduation Rates, RMSPE, and p -value.

Specification	Year 1	Year 2	Year 3	RMSPE	p -value
Panel A: Restricted donor pool					
Specification 1	-3.373	-1.015	-0.469	2.53	.696
Specification 2	-9.611	-1.763	-3.344	3.14	.435
Specification 3	-9.255	-8.684	-6.442	6.88	.348
Specification 4	-6.076	-0.908	-1.911	2.21	.348
Specification 5	-7.505	-0.837	0.842	1.95	.565
Specification 6	-8.062	-2.268	-3.646	3.26	.174
Panel B: Comprehensive donor pool					
Specification 1	-8.753	-1.421	-1.071	2.52	.101
Specification 2	-13.274	-6.286	0.367	2.84	.055
Specification 3	-11.527	-8.127	-3.870	5.79	.215
Specification 4	-8.904	-2.038	-1.913	2.66	.106
Specification 5	-6.917	-6.630	6.850	2.48	.557
Specification 6	-13.202	-8.398	-2.984	3.06	.076

Note. Restricted donor pool includes 22 districts, whereas comprehensive donor pool includes 236 districts. Years 1–3 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 Table 1. Pretreatment period includes years 2001–2007. RMSPE = the root mean squared prediction error.

Conclusions

This article uses synthetic control methods to evaluate the effects of a place-based scholarship program and to highlight critical issues evaluators face in using the technique to construct valid counterfactual estimates. In the analysis of enrollments, where the time series of outcome measures does not appear to be strongly influenced by transitory shocks, synthetic control methods provide consistent evidence of meaningful public school enrollment increases in Syracuse relative to control districts after the start of the Say Yes to Education program. Moreover, these effect estimates did not vary substantially when choices of donor pool and pretreatment outcome years were changed, providing an important check on the robustness of the results. Because the synthetic control method disciplines the choice of comparison units used to estimate counterfactual posttreatment outcomes, we are able to derive effect estimates that are largely insensitive to the sample of units used in the estimation and remove a potentially important source of bias in effect estimates.

In contrast to enrollments, the time series of graduation rates is quite volatile. In this case, the effect estimates provided by synthetic control methods were sensitive to both the choice of pretreatment years included in the matching algorithm and the choice of donor pool districts. Given these results, and that synthetic controls can yield minimally biased effect estimates only when the number of preintervention time periods is large relative to the scale of transitory shocks (Abadie et al., 2010), it is difficult to draw conclusions about the effects of the treatment on graduation rates. It is important to note, however, that estimates derived from parametric regressions can also be sensitive to both functional form and sample choices when pretreatment outcomes trends are noisy.

More substantively, it is quite possible that meaningful effects on graduation rates in Syracuse will take time to develop. Students already at risk of dropping out when the program was announced may be unmotivated by the scholarship offer or, more importantly, may already be too far behind to graduate. As students who received additional supports in elementary and middle school begin to reach graduation age, it is possible that graduation rates may start to increase.

Finally, we find consistent evidence that, after decades of population and enrollment declines, public school enrollments in the city of Syracuse began to rebound following the start of the Say Yes to Education program. While the estimated 4-year increases of between 3% and 6% may appear modest, it

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 `tset unit time`.

```
synth depvar predictorvars, options
```

Gives you:

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

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 (specifically, Black males)? 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?

- 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`