Synthetic control

A. Colin Cameron
Univ. of California, Davis

These slides are part of the set of slides
A. Colin Cameron, Introduction to Causal Methods
https://cameron.econ.ucdavis.edu/causal/

March 2023

Introduction

- These slides give an introductory example of synthetic control
    - synthetic control is a method for causal inference
    - it is method for when a policy affects only one group out of many groups
    - it compares the outcome for the treated group to a synthetic control that is an appropriately weighted average of the untreated groups.
- The method relies on creating an appropriate synthetic control
    - and inference methods for synthetic control are still an active area of research.

- Separately the Stata file `dsynth.do` implements these methods
  - using dataset AED_SMOKING.DTA
- The data are from chapter 25.6.2 of A. Colin Cameron and Pravin K. Trivedi (2022), *Microeconometrics using Stata,* Volume 2.
- The data source is Galiani and Quistorff (2017), "The synthrunnerpackage: Utilities to automate synthetic control estimation using synth," *Stata Journal*, 834-849.
  - based in turn on Abadie, Diamond and Hainmuller (2010), JASA, 493-505.

Outline

Synthetic control

- We have data on a single treated group and several untreated groups. This data is over time both before and after the treatment period.
- Let y denote the outcome.
- Let s denote the groups (e.g. state)
  - $s = 1$ for the treated state and $s = 2, ..., J + 1$ for the J untreated.
- Treatment occurs between T_0 and T_1
  - we have pre-treatment data $t = 1, ..., T_0$
  - and post-treatment data $t = T_1, ..., T$.

Synthetic control (continued)

- We construct time-invariant weights w_s (a function of $x's$) so that in the pre-period
  - treated $y_{1t} \simeq \frac{1}{J}\sum_{s=2}^{J+1} w_s y_{st}$ for each $t = 1, ..., T_0$
  - the weights use data provided on key control variables likely to determine y
  - the original 2010 article by Abadie et al. provides an algorithm.
  - usually $w_s = 0$ for all but a few untreated groups.
- The synthetic control treatment effect uses the same weights in the post-treatment periods to give
  - Treatment effect $= y_{1t} - \frac{1}{J}\sum_{s=2}^{J+1} w_s y_{st}$ for $t = T_1, ..., T$.

Example: Tobacco control in California

- Does tobacco control lead to reduced smoking?
- Proposition 99 implemented in 1989 increased California's cigarette excise tax by a large amount per pack and required that this money to be used for anti-smoking education and media campaigns and for local clean indoor-air ordinances throughout the state.
- Annual data for California and 38 untreated states from 1970 to 2000.
- Outcome $y =$ `cigsale` is annual cigarette packs sold per capita.
- Variables used to compute the weights w_s are per capita beer consumption and log income, retail price of a cigarette pack and percentage of population aged 15-24 years.

Example (continued)

- Summary statistics for key variables

```
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
state           long    %14.0g     state      State no
year            float   %9.0g                 Year
cigsale         float   %9.0g                 Cigarette sale per capita (in
                                                packs)
lnincome        float   %9.0g                 Log state per capita GDP
beer            float   %9.0g                 Beer consumption per capita
age15to24       float   %9.0g                 Percent of state population aged
                                                15-24 years
retprice        float   %9.0g                 Retail price of cigarettes
-------------------------------------------------------------------------------
Sorted by: year

. summarize, sep(7)

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       state |      1,209          20    11.25929          1         39
        year |      1,209        1985    8.947973       1970       2000
     cigsale |      1,209    118.8932     32.7674       40.7      296.2
    lnincome |      1,014    9.861634    .1706769   9.397449   10.48662
        beer |        546     23.4304     4.22319        2.5       40.4
   age15to24 |        819     .175472    .0151589   .1294482   .2036753
    retprice |      1,209    108.3419    64.38199       27.3      351.2
```

Results

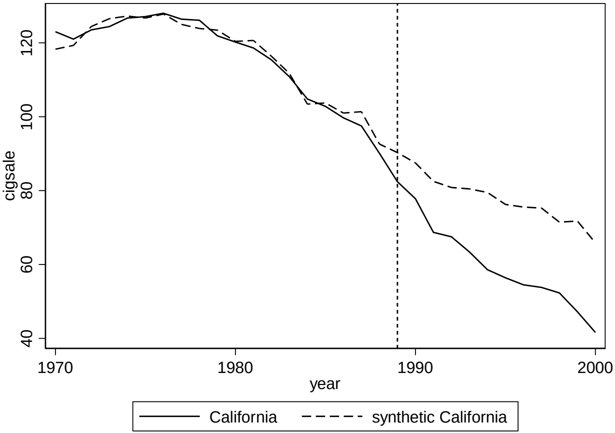
- The treatment period is 1989 and California is the third state in the dataset.
- The weights use average beer sales over 1984-88, average log income over 1972-1988, averages over 1970-1988 in retail price and percentage aged 15-24, and cigarette sales in three selected years.
- We use the user-written Stata command `synth`

```
* Synthetic control: synth command
tsset state year
synth cigsale beer(1984(1)1988) lnincome(1972(1)1988) ///
retprice age15to24 cigsale(1988) cigsale(1980) ///
cigsale(1975), trunit(3) trperiod(1989) figure
```

- The weights are all zero except for Colorado (0.285), Connecticut (0.101), Nevada (0.245) and Utah (0.369).

Results: Graph

- This graph plots the outcome for California against that for the synthetic control in pre- and post-treatment periods.
- Treatment effect (post-treatment difference in the curves) is large.



Results: Numbers

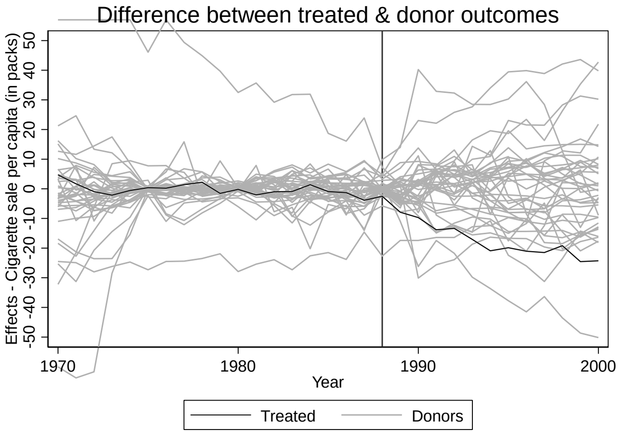
- The effects each year are very large compared to the sample mean.
- The `pvals` column gives the fraction of times that the treatment effect is larger for California than if the same analysis is repeated for each of the other states (with California omitted).
  - we then expect no treatment effect as they weren't treated
  - e.g. C12 is 2000 with 4 having $|TE| > 24.2687$ and $4/38 = 0.105263$.

Post-treatment results: Effects, p-values, standardized p-values

	estimates	pvals	pvals_std
c1	-7.887098	.1315789	0
c2	-9.693599	.1842105	0
c3	-13.8027	.2105263	0
c4	-13.344	.1315789	0
c5	-17.0624	.1052632	0
c6	-20.8943	.0789474	0
c7	-19.8568	.1315789	.0263158
c8	-21.0405	.1578947	0
c9	-21.4914	.1052632	.0263158
c10	-19.1642	.1842105	.0263158
c11	-24.554	.1052632	0
c12	-24.2687	.1052632	.0263158

Results: California and Placebo treatment effects

- This graph gives the treatment effect for California (bold line) and each of the other states (for whom we expect no treatment effect).



Further Details

- Better methods to obtain a p-value are currently being developed.

References

- Alberto Abadie (2021), "Using synthetic controls: Feasibility, data requirements, and methodological aspects," *Journal of Economic Literature*, 59: 391-425.
- These books are given in approximate order of increasing difficulty.
- Cunningham, Scott (2021), Causal Inference: The MixTape, Yale University Press, chapter 10.
- A. Colin Cameron and Pravin K. Trivedi (2022), Microeconometrics using Stata: Volume 2, Second Edition, Stata Press, chapter 25.6.