Synthetic Controls

Alexander Torgovitsky

ECON 31720: Applied Microeconometrics
University of Chicago, Fall 2019

Overview — Abadie & Gardeazabal (2003), Abadie et al (2010, 2015)

- Difference-in-difference with one treated unit (a “case study”)
- Suppose common trends is unappealing for all untreated units
- Construct a weighted average of untreated units to use for counterfactuals
- Weighted average is chosen to match pre-period outcomes/covariates

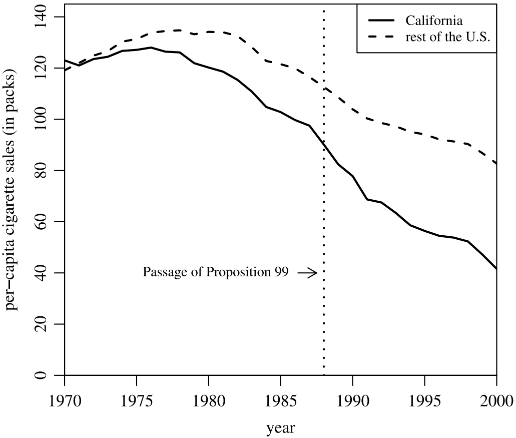
→ Selection-on-observables, but weighted in a particular way

- Several important formal questions about the procedure remain open

Empirical example (Abadie, Diamond, & Hainmueller 2010)

- What was the effect of California’s anti-tobacco law on cigarette sales?
- Proposition 99, passed in 1988, one of the first (modern) laws of its kind
- California already had sharply declining smoking rates before 1988
- Probably one reason the law was passed — common trends unlikely
- No other US state had the same trend — so combine several of them

Cigarette Sales Pre-Trends



- Divergent pre-period trends between California and rest of U.S.
- Possible that other *individual* states had similar trends (but they don't)

Notation

- Let X_1 denoted treated group pre-period outcomes and covariates
- Let X_0 denote untreated group pre-period outcomes and covariates
- k variables and J untreated groups, so X_1 is $k \times 1$ and X_0 is $k \times J$
- Treated outcomes $Y_{t,1}$ and untreated outcomes $Y_{t,0}$ (a $J \times 1$ vector)

Solve the following optimization problem to create weights

$$w^\star \equiv \underset{w \in \mathbb{R}^J}{\arg\min} \|X_1 - X_0 w\|_V \quad \text{s.t.} \quad w_j \geq 0 \; \forall j \quad \text{and} \quad \sum_{j=1}^{J} w_j = 1$$

- "Synthetic California" is the $w^\star$–linear combination of untreated units
- V is a weighting matrix (let's call it a norm matrix) — more on this later
- Weights are chosen to get the best (convex) fit in the pre-period
- Treated effect estimate for the post period is $Y_{t,1} - Y_{t,0}' w^\star$

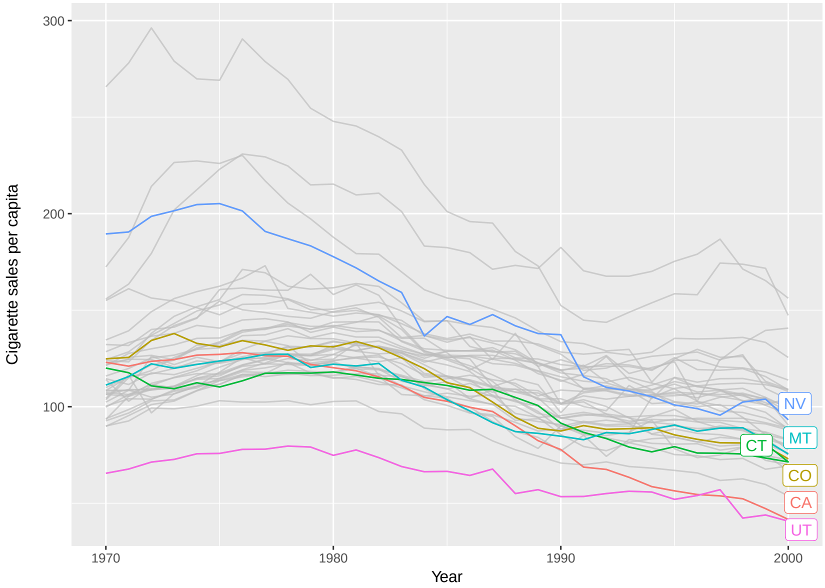
Synthetic California

Table 2. State weights in the synthetic California

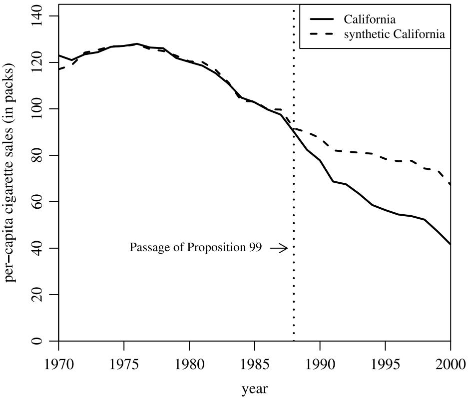
State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	–	Nebraska	0
Arizona	–	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	–
Connecticut	0.069	New Mexico	0
Delaware	0	New York	–
District of Columbia	–	North Carolina	0
Florida	–	North Dakota	0
Georgia	0	Ohio	0
Hawaii	–	Oklahoma	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

- — states were removed because they passed similar laws in post-period
- Solution is typically sparse as here: NV, UT, MT, CO and a bit of CT

Series Used in Synthetic California



Synthetic California Pre-Trends



Variables used to construct Synthetic California

Lagged outcome (per capita sales) in 1975, 1980, 1988

Cigarette price, log per capita income, percentage aged 15-24, per capita beer consumption (all averaged over 1980–1988)

Estimated Treatment Effects

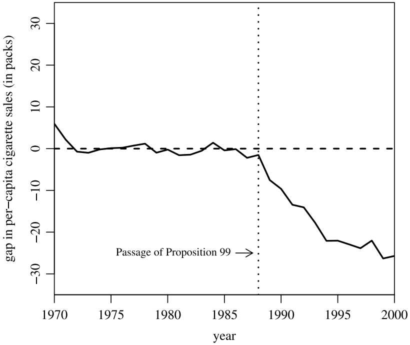


Figure 3. Per-capita cigarette sales gap between California and synthetic California.

- Subtract synthetic California from real California in post-period
- 25% reduction (20 packs) averaged over the entire post-period
- Considerably larger estimate than in previous papers

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

NOTE: All variables except lagged cigarette sales are averaged for the 1980–1988 period (beer consumption is averaged 1984–1988). GDP per capita is measured in 1997 dollars, retail prices are measured in cents, beer consumption is measured in gallons, and cigarette sales are measured in packs.

- Matched covariates for real and synthetic are similar — *by construction*
- Of course what we want is untreated post-period series to be similar
- Under what assumptions does the procedure ensure this?

Take a step back

- The motivation of synthetic controls is weakening common trends

$\rightarrow$ Otherwise one could just use standard DID methods

- However clearly we need a model to justify the weighting procedure

$\rightarrow$ Matching on pre-periods need not say *anything* about the post-period

A factor model for untreated outcomes

- Consider a simple case with $J = 2$ untreated, $t = 1, 2$ pre-periods

$$Y_{it}(0) = A_i + B_t + L_i F_t + V_{it} \quad \text{for } t = 1, 2 \text{ and } i = 0, 1, 2$$

- Suppose we perfectly match treated (Y_{01}, Y_{02}) with $\{Y_{i1}, Y_{i2}\}_{i=1}^{2}$
- Then we (almost) matched the factors (see supplement) — use to forecast
- Some rank conditions are required — these may be hard to interpret
- Intuitively, there should be "few" factors relative to pre-period length
- Also, dealing with the idiosyncratic part formally is a bit unclear

Choosing the Norm Matrix

Criteria

- Abadie et al (2010) suggest the pre-period fit of the synthetic control:

$$V^{\star} \equiv \arg\min_{V} \sum_{t=1}^{t_0} \left(Y_{t,1} - Y_{t,0}' w^{\star}(V)\right)^2$$

- Abadie et al (2015) do the same but with training/validation split

→ Mitigates potential overfitting from using $Y_{t,0}$ twice

Complications

- Estimating synthetic control for fixed V was a simple quadratic program
- Solving for $V^{\star}$ is a much harder, less structured problem

→ $w^{\star}(V)$ is certainly not a linear function of V (maybe not even smooth)

- Moreover, V is a $k \times k$ matrix — relatively high-dimensional
- Implemented in the `synth` R package, but I have heard unreliably
- A safer choice may be some sort of Mahalonobis distance

Uniqueness of Weights

The fit of the synthetic control

- Proponents of synthetic control say don't use the method if fit is "bad"

→ How large does this need to be to be "bad"? Hard to say …

- $\|X_1 - X_0 w^\star\|_V = 0$ (fit is perfect) $\quad\Leftrightarrow\quad$ X_1 is in the convex hull of X_0
- Synthetic California was almost in the convex hull — see balance table

But very fit also creates problems

- $w^\star$ is (generically) unique if *and only if* X_1 is *not* in the convex hull of X_0

→ That is, only if you *can't* get a perfect fit

- More likely with fewer k and fewer untreated units
- If a perfect fit is possible, then (typically) will be infinite equally good $w^\star$

→ Problem is that does not imply $Y_{t,0}' w^\star$ is the same for all such $w^\star$!

- Abadie & L'Hour (2019) propose a penalization based on $\|X_1 - X_0\|$

→ Can restore uniqueness, but adds another tuning parameter

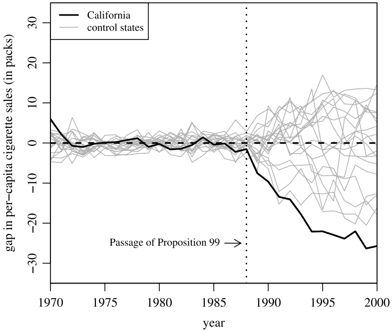


Figure 7. Per-capita cigarette sales gaps in California and placebo gaps in 19 control states (discards states with pre-Proposition 99 MSPE two times higher than California's).

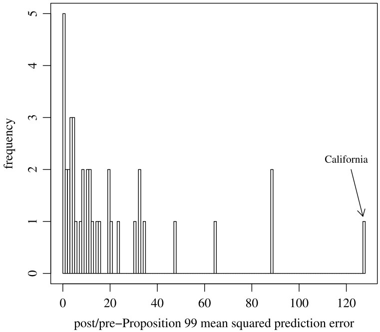


Figure 8. Ratio of post-Proposition 99 MSPE and pre-Proposition 99 MSPE: California and 38 control states.

- Difficult setting with one treated group, *plus* complication of weights
- Some recent proposals, but still unclear how to do inference formally
- Abadie et al (2010) use a placebo exercise (randomization inference)
- Measure in two ways: Treatment effect and ratio of pre-post MSE

Why not just use a factor model?

- There is a large and old literature on estimating factor models
- Why not estimate one directly? (Gobillon and Magnac 2016, Xu 2017)
- More direct than fitting weights and justifying ex-post by a factor model
- Possible response is not needing to choose the exact number of factors

Selection on observables

- Synthetic control is a way of implementing selection on observables

→ Already have many ways — the *estimator* was not the problem

- Doesn't do anything to address selection on *unobservables*

Statistical inference

- Major conceptual difficulties with statistical inference (not all new)
- Really a problem with case studies — not having large N and/or T
- Active line of research in thinking about procedures in various settings