Treatment Effects
Part 2

Richard L. Sweeney

based on slides by Chris Conlon

Empirical Methods
Spring 2021

Overview

This lecture draws heavily upon

- Pamela Jakiela and Owen Ozier's slides.)
- The Mixtape (read online here)

Difference in Differences

- Sometimes we may feel we can impose more structure on the problem.
- Suppose in particular that we can write the outcome equation as

$$Y_{it} = \alpha_i + d_t + \beta_i T_{it} + u_{it}$$

- In the above we have now introduced a time dimension $t = \{1, 2\}$.
- Now suppose that $T_{i1} = 0$ for all i and $T_{i2} = 1$ for a well defined group of individuals in our population.
- This framework allows us to identify the ATT effect under the assumption that the growth of the outcome in the non-treatment state is independent of treatment allocation:

$$E[Y^0_{i2} - Y^0_{i1} | T] = E[Y^0_{i2} - Y^0_{i1}]$$

- This is known as parallel trends.

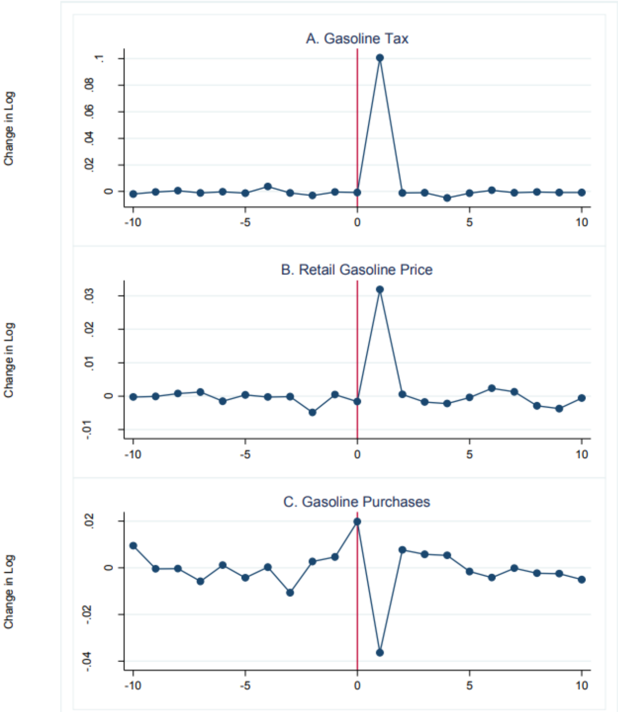
Before and After

An even simpler estimator is the event study.

- We look an outcome before or after an event
  - A news event: the announcement of a merger or stock split.
  - A tax change, a new law, etc.

$$\begin{aligned} E[Y_{i2} - Y_{i1} | T_{i2} = 1] &= E[Y^1_{i2} - Y^1_{i1} | T_{i2} = 1] \\ &= d_2 - d_1 + E[\beta_i | T_{i2} = 1] \end{aligned}$$

- Except under strong conditions $d_2 = d_1$ we shouldn't believe the results of the before and after estimator.
- Main Problem: we attribute changes to treatment that might have happened anyway trend.
- e.g: Cigarette consumption drops 4% after a tax hike. (But it dropped 3% the previous four years).
- Also worry about: anticipation, gradual rollout, etc.


A. Gasoline Tax
Change in Log
0
.02
.04
.06
.08
.1
-10
-5
0
5
10
B. Retail Gasoline Price
Change in Log
-.01
0
.01
.02
.03
-10
-5
0
5
10
C. Gasoline Purchases
Change in Log
-.04
-.02
0
.02
-10
-5
0
5
10

Difference in Differences

Let's try and estimate $d_2 - d_1$ directly and then difference it out. Here we use parallel trends:

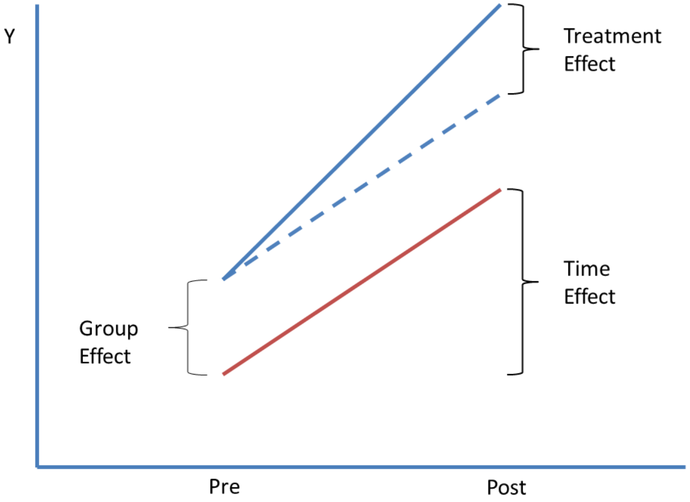
$$E[Y_{i2}^0 - Y_{i1}^0 | T_{i2} = 1] = E[Y_{i2}^0 - Y_{i1}^0 | T_{i2} = 0]$$
$$E[Y_{i2} - Y_{i1} | T_{i2} = 0] = d_2 - d_1$$

We now obtain an estimator for ATT:

$$E[\beta_i | T_{i2} = 1] = E[Y_{i2} - Y_{i1} | T_{i2} = 1] - E[Y_{i2} - Y_{i1} | T_{i2} = 0]$$

which can be estimated by the difference in the growth between the treatment and the control group.

Parallel trends solves a "missing data" problem



Example: Minimum Wage

Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania

By DAVID CARD AND ALAN B. KRUEGER*

On April 1, 1992, New Jersey's minimum wage rose from $4.25 to $5.05 per hour. To evaluate the impact of the law we surveyed 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. Comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant) provide simple estimates of the effect of the higher minimum wage. We also compare employment changes at stores in New Jersey that were initially paying high wages (above $5) to the changes at lower-wage stores. We find no indication that the rise in the minimum wage reduced employment. (*JEL* J30, J23)

Example: Minimum Wage

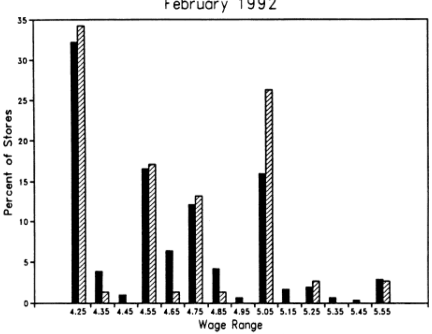
TABLE 1—SAMPLE DESIGN AND RESPONSE RATES

		Stores in:	
	All	NJ	PA
Wave 1, February 15 – March 4, 1992:			
Number of stores in sample frame:[a]	473	364	109
Number of refusals:	63	33	30
Number interviewed:	410	331	79
Response rate (percentage):	86.7	90.9	72.5
Wave 2, November 5 – December 31, 1992:			
Number of stores in sample frame:	410	331	79
Number closed:	6	5	1
Number under rennovation:	2	2	0
Number temporarily closed:[b]	2	2	0
Number of refusals:	1	1	0
Number interviewed:[c]	399	321	78

[a]Stores with working phone numbers only; 29 stores in original sample frame had disconnected phone numbers.

[b]Includes one store closed because of highway construction and one store closed because of a fire.

[c]Includes 371 phone interviews and 28 personal interviews of stores that refused an initial request for a phone interview.


February 1992
Percent of Stores
35
30
25
20
15
10
5
0
4.25 4.35 4.45 4.55 4.65 4.75 4.85 4.95 5.05 5.15 5.25 5.35 5.45 5.55
Wage Range

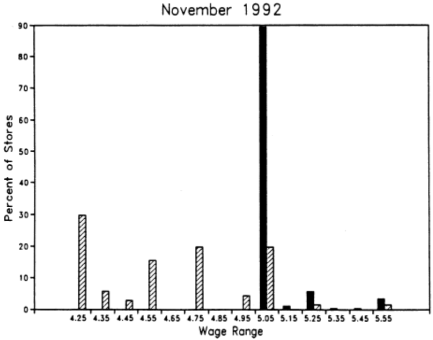

November 1992
Percent of Stores
90
80
70
60
50
40
30
20
10
0
4.25 4.35 4.45 4.55 4.65 4.75 4.85 4.95 5.05 5.15 5.25 5.35 5.45 5.55
Wage Range

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

	Stores by state			Stores in New Jersey[a]			Differences within NJ[b]	
Variable	PA (i)	NJ (ii)	Difference, NJ – PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	–2.69 (1.37)	–2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	–2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores[c]	–2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	–2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d]	–2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	–2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

Notes: Standard errors are shown in parentheses. The sample consists of all stores with available data on employment. FTE (full-time-equivalent) employment counts each part-time worker as half a full-time worker. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing.

[a]Stores in New Jersey were classified by whether starting wage in wave 1 equals \$4.25 per hour ($N = 101$), is between \$4.26 and \$4.99 per hour ($N = 140$), or is \$5.00 per hour or higher ($N = 73$).

[b]Difference in employment between low-wage (\$4.25 per hour) and high-wage (≥ \$5.00 per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

[c]Subset of stores with available employment data in wave 1 and wave 2.

[d]In this row only, wave-2 employment at four temporarily closed stores is set to 0. Employment changes are based on the subset of stores with available employment data in wave 1 and wave 2.

Difference in Differences

Now consider the following problem:

- Suppose we wish to evaluate a training program for those with low earnings. Let the threshold for eligibility be B.
- We have a panel of individuals and those with low earnings qualify for training, forming the treatment group.
- Those with higher earnings form the control group.
- Now the low earning group is low for two reasons
  1. They have low permanent earnings (α_i is low) - this is accounted for by diff in diffs.
  2. They have a negative transitory shock (u_{i1} is low) - this is not accounted for by diff in diffs.
- #2 above violates the assumption $E[Y^0_{i2} - Y^0_{i1}|T] = E[Y^0_{i2} - Y^0_{i1}]$.
- This is effectively regression to the mean: those unlucky enough to have a bad shock recover and show greater growth relative to those with a good shock. The nature of the bias depends on the stochastic properties of the shocks and how individuals select into training.

Who get's treated?

- The assumption on growth of the non-treatment outcome being independent of assignment to treatment may be violated, but it may still be true conditional on X.
- Consider the assumption

$$E[Y_{i2}^0 - Y_{i1}^0 | X, T] = E[Y_{i2}^0 - Y_{i1}^0 | X]$$

- This is just matching assumption on a redefined variable, namely the growth in the outcomes. In its simplest form the approach is implemented by running the regression

$$Y_{it} = \alpha_i + d_t + \beta_i T_{it} + \gamma_t' X_i + u_{it}$$

which allows for differential trends in the non-treatment growth depending on X_i. More generally one can implement propensity score matching on the growth of outcome variable when panel data is available.

DiD with Repeated Cross Sections

- Suppose we do not have available panel data but just a random sample from the relevant population in a pre-treatment and a post-treatment period.
- First consider a simple case where $E[Y^0_{i2} - Y^0_{i1}|T] = E[Y^0_{i2} - Y^0_{i1}]$.
- We need to modify slightly the assumption to

$$E[Y^0_{i2}|_{\text{Group receiving training}}] - E[Y^0_{i1}|_{\text{Group receiving training in the next period}}] = E[Y^0_{i2} - Y^0_{i1}]$$

  which requires additional assumption that the population we will be sampling from does not change composition.
- We can then obtain immediately an estimator for ATT as

$$E[\beta_i|T_{i2} = 1]$$
$$= E[Y_{i2}|_{\text{Group receiving training}}] - E[Y_{i1}|_{\text{Group receiving training next period}}]$$
$$- \{E[Y_{i2}|_{\text{Non-trainees}}] - E[Y_{i1}|_{\text{Group not receiving training next period}}]\}$$

Difference in Differences with Repeated Cross Sections

- More generally we need an assumption of conditional independence of the form

$$E[Y_{i2}^0|X, \text{Group receiving training}] - E[Y_{i1}^0|X, \text{Group receiving training next period}]$$
$$= E[Y_{i2}^0|X] - E[Y_{i1}^0|X]$$

- Under this assumption (and some auxiliary parametric assumptions) we can obtain an estimate of the effect of treatment on the treated by the regression

$$Y_{it} = \alpha_g + d_t + \beta T_{it} + \gamma' X_{it} + u_{it}$$

Difference in Differences with Repeated Cross Sections

- More generally we can first run the regression

$$Y_{it} = \alpha_g + d_t + \beta(X_{it})T_{it} + \gamma' X_{it} + u_{it}$$

  where α_g is a dummy for the treatment of comparison group, and $\beta(X_{it})$ can be parameterized as $\beta(X_{it}) = \beta' X_{it}$. The ATT can then be estimated as the average of $\beta' X_{it}$ over the (empirical) distribution of X.
- A non parametric alternative is offered by Blundell, Dias, Meghir and van Reenen (2004).

DiD vs Fixed Effects

- What if we have a long panel with many similar changes?
    - Greenstone (2002): Counties move in and out of Clean Air Act
    - Evans, Ringel, and Stech (1999): Since 1975, more than 200 state cigarette tax changes
- Fixed effects generalize DD with $T > 2$ periods and $J > 2$ groups
- Advantage relative to DD: more precise estimates by pooling several changes
- Disadvantage: fixed effects is a black-box regression, more difficult to check trends non-parametrically as with a single change

The best DiD's can be seen graphically

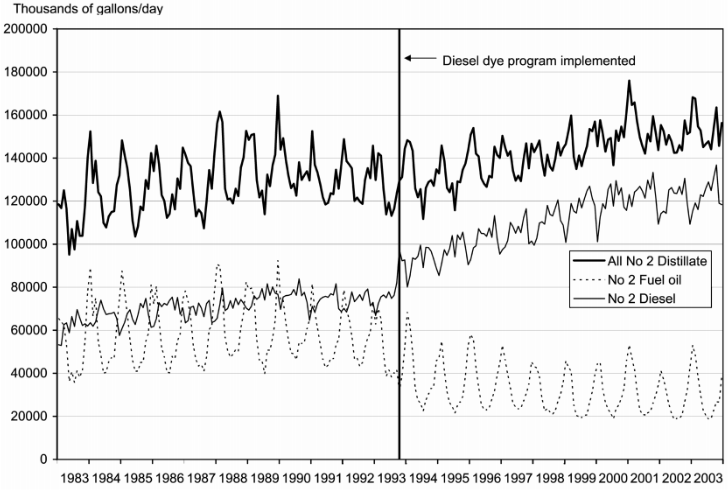


FIG. 2.—U.S. sales of No. 2 distillate

What about triple differencing?

- Sometime we might use a "placebo" DD to make parallel trends more convincing
- Example: Imagine a policy which offered STEM outreach to high school girls in Massachusetts
  - Natural DiD control group: boys in MA
  - However over time there could be general shifts in the relative outcomes of boys and girls everywhere
  - Suggest looking at how the difference between boys and girls in MA changed relative to the changes in other states (say RI)
- Logically sound, but much harder to see/ validate visually

Difference in Differences and Selection on Unobservables

- Suppose we relax the assumption of *no selection* on unobservables.
- Instead we can start by assuming that

$$E[Y_{i2}^0|X,Z] - E[Y_{i1}^0|X,Z] = E[Y_{i2}^0|X] - E[Y_{i1}^0|X]$$

  where Z is an instrument which determines training eligibility say but does not determine outcomes in the non-training state. Take Z as binary (1,0).
- Non-Compliance: not all members of the eligible group $(Z = 1)$ will take up training and some of those ineligible $(Z = 0)$ may obtain training by other means.
- A difference in differences approach based on grouping by Z will estimate the impact of being allocated to the eligible group, but not the impact of training itself.

Difference in Differences and Selection on Unobservables

- Now suppose we still wish to estimate the impact of training on those being trained (rather than just the effect of being eligible)
- This becomes an IV problem and following up from the discussion of LATE we need stronger assumptions
  - Independence: for $Z = a, \{Y^0_{i2} - Y^0_{i1}, Y^1_{i2} - Y^1_{i1}, T(Z = a)\}$ is independent of Z.
  - Monotonicity $T_i(1) \geq T_i(0) \, \forall \, i$
- In this case LATE is defined by

$$[E(\Delta Y|Z = 1) - E(\Delta Y|Z = 0)]/[Pr(T(1) = 1) - Pr(T(0)$$

assuming that the probability of training in the first period is zero.

Synthetic Controls

- DiD methods compare two groups before and after some change.
- Challenge: What's a good comparison group? Even if you pick the best available option, might not track eachother that closely even in the pre-period.
- Moreover, if we don't have another untreated group that is well balanced against the treatment group, are we stuck?
- Synthetic control methods pick weighted averages from control population to construct better comparisons (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010)
- Athey and Imbens (2017) call this "arguably the most important innovation in the policy evaluation literature in the past 15 years".

Initial motivation: Case studies

- Often we're interested in the aggregate effects of large, singular policies.
  - What was the impact of MassHealth?
  - Fukushima
  - Terrorism
  - German Re-unification
- What would a rigorous "case study" of these look like?

ADH (JASA 2010)

- Consider a panel with $J+1$ units observed for $t = 1, 2, ..., T$ periods.
- Unit 1 exposed to treatment in period T_0 (continues to T)
- Synthetic control estimator is

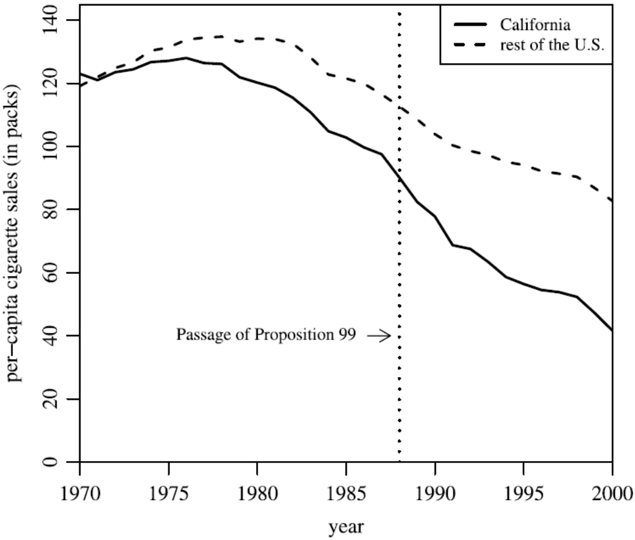
$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

  where w is a collection of weights.
- In Abadie, Diamond, and Hainmueller (2010) the (non-negative) weights are chosen to minimize the distance between some chosen vector of preintervention characteristics (and sum to one).
- Subsequent literature has relaxed these.

ADH Example: CA Prop 99

- Anti cigarette law in CA in 1988
  - increased state excise tax by 25 cents per pack
  - earmarked the tax revenues to health and anti-smoking education budgets
  - funded anti-smoking media campaigns
  - spurred local clean indoor-air ordinances throughout the state
- What was the net effect on sales?

Sales were trending down everywhere



What does synthetic CA look like?

Table 2. State weights in the synthetic California

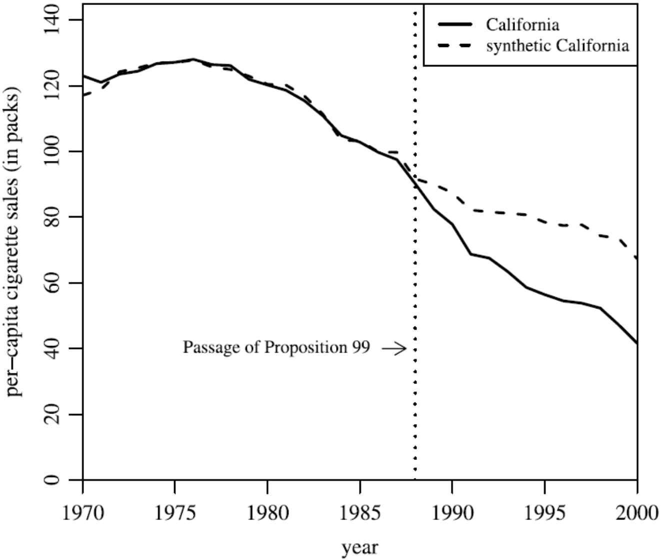
State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	–	Nebraska	0
Arizona	–	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	–
Connecticut	0.069	New Mexico	0
Delaware	0	New York	–
District of Columbia	–	North Carolina	0
Florida	–	North Dakota	0
Georgia	0	Ohio	0
Hawaii	–	Oklahoma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

Balance Acheived

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.


California
synthetic California
Passage of Proposition 99
per–capita cigarette sales (in packs)
year
0
20
40
60
80
100
120
140
1970
1975
1980
1985
1990
1995
2000

Parallel trends acheived by contstruction

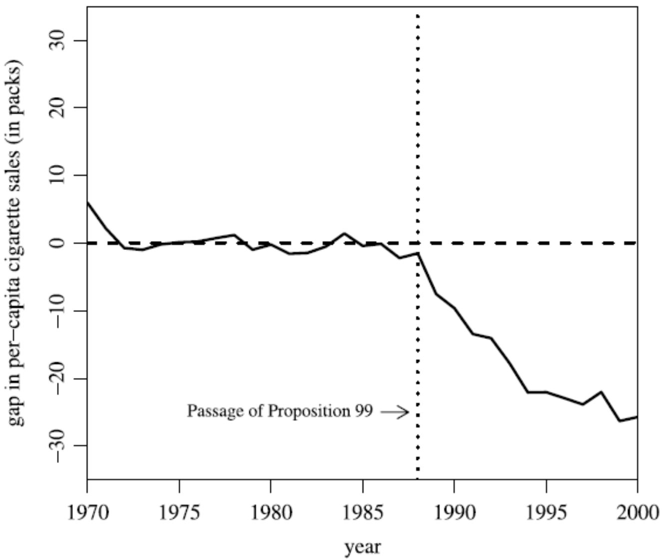


Figure 3. Per-capita cigarette sales gap between California and syn-

What about inference

- SE's typically reported reflect uncertainty in sample relative to aggregate population.
- ADH propose using a placebo test to assess null of no change in CA.
- Steps:
  1. Randomly select one of the other J control units / time cutoffs and declare it treated.
  2. Construct synthetic controls and estimate ATT.
  3. Repeat many times
- Since none of these units are actually treated, this test distribution simulates distribution of the differences relative to the synthetic control under the true null of no effect.

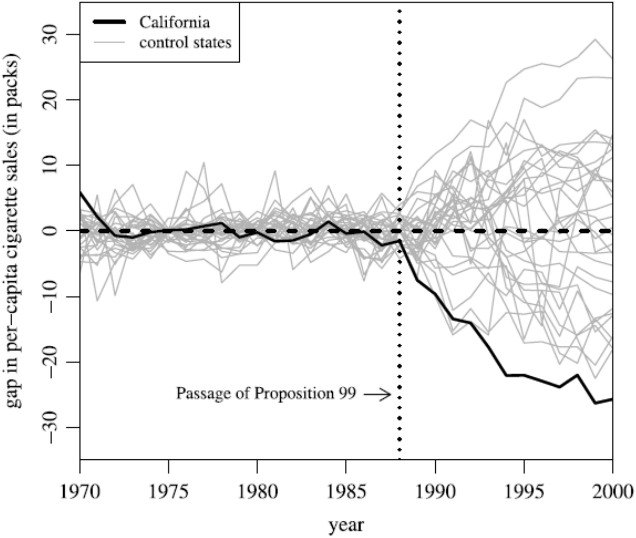


Figure 6. Per-capita cigarette sales gaps in California and placebo gaps in 29 control states (discards states with pre-Proposition 99 MSPE five times higher than California's).

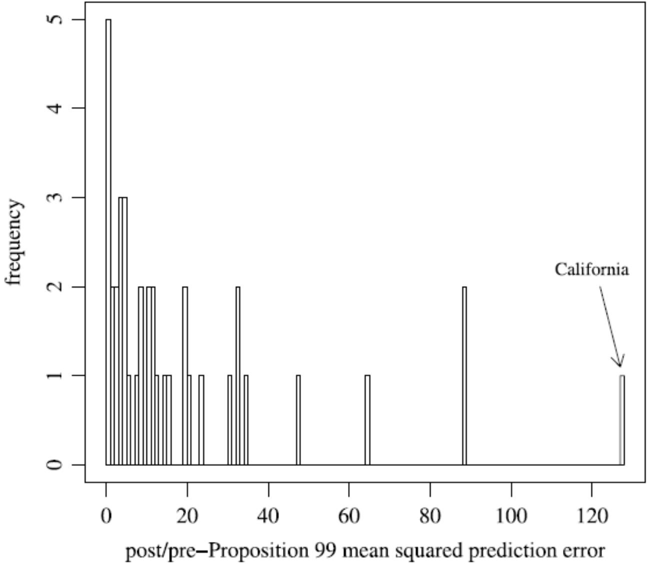


Figure 8. Ratio of post-Proposition 99 MSPE and pre-Proposition 99 MSPE: California and 38 control states.

Synthetic Difference in Differences

Dmitry Arkhangelsky[†] Susan Athey[‡] David A. Hirshberg[§]

Guido W. Imbens[¶] Stefan Wager[‖]

Draft version November 2020

Abstract

We present a new estimator for causal effects with panel data that builds on insights behind the widely used difference in differences and synthetic control methods. Relative to these methods, we find, both theoretically and empirically, that the proposed "synthetic difference in differences" estimator has desirable robustness properties, and that it performs well in settings where the conventional estimators are commonly used in practice. We study the asymptotic behavior of the estimator when the systematic part of the outcome model includes latent unit factors interacted with latent time factors, and we present conditions for consistency and asymptotic normality.

Arkhangelsky et al. (2020) synthesize recent developments in synthetic controls, DiD, regularization. Video lecture available here.

while the last $N_{\text{tr}} = N - N_{\text{co}}$ (treated) units are exposed after time T_{pre}. Similar to SC, we start by finding weights $\hat{\omega}^{\text{sdid}}$ that align pre-exposure trends in the outcome of unexposed units with those for the exposed units; $\sum_{i=1}^{N_{\text{co}}} \hat{\omega}_i^{\text{sdid}} Y_{it} \approx N_{\text{tr}}^{-1} \sum_{i=N_{\text{co}}+1}^{N} Y_{it}$ for all $t = 1, \ldots, T_{\text{pre}}$. We also find time weights $\hat{\lambda}_t^{\text{sdid}}$ that similarly balance pre-exposure time periods with post-exposure ones (see Section 2 for details). Then we use these weights in a basic two-way fixed effects regression to estimate the causal effect of exposure (denoted by τ):[1]

$$\left(\hat{\tau}^{\text{sdid}}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \underset{\tau,\mu,\alpha,\beta}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau \right)^2 \hat{\omega}_i^{\text{sdid}} \hat{\lambda}_t^{\text{sdid}} \right\}. \tag{1.1}$$

In comparison, DID estimates the effect of treatment exposure by solving the same two-way fixed effects regression problem without either time or unit weights:

$$\left(\hat{\tau}^{\text{did}}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \underset{\alpha,\beta,\mu,\tau}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau \right)^2 \right\}. \tag{1.2}$$

Algorithm 1: Synthetic Difference in Differences (SDID)

Data: $\boldsymbol{Y}, \boldsymbol{W}$

Result: Point estimate $\hat{\tau}^{\text{sdid}}$

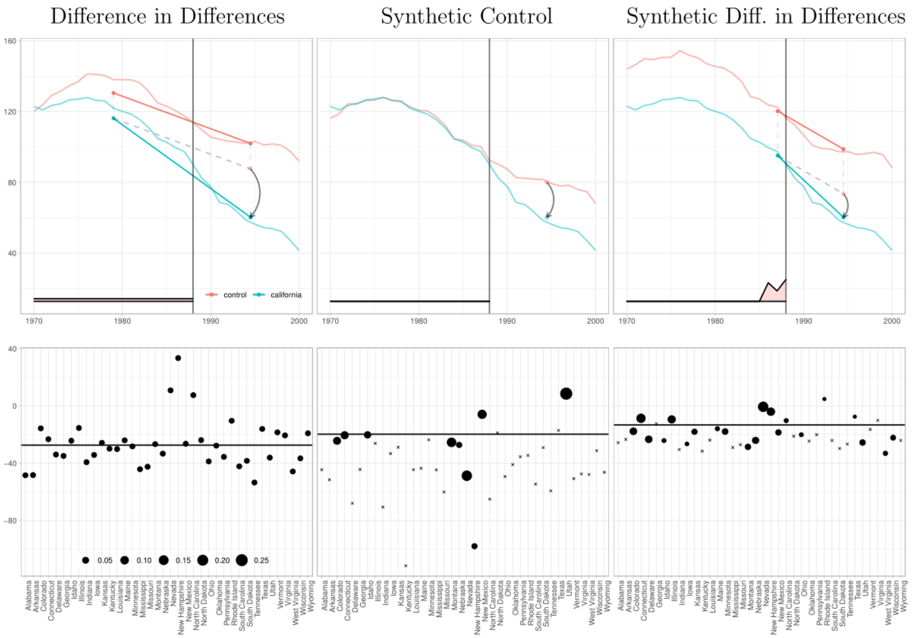
Compute regularization parameter ζ using (2.2);

Compute unit weights $\hat{\omega}^{\text{sdid}}$ via (2.1);

Compute time weights $\hat{\lambda}^{\text{sdid}}$ via (2.3);

Compute the SDID estimator via the weighted DID regression

$$\left(\hat{\tau}^{\text{sdid}}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \underset{\tau,\mu,\alpha,\beta}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau\right)^2 \hat{\omega}_i^{\text{sdid}} \hat{\lambda}_t^{\text{sdid}} \right\};$$



Difference in Differences
Synthetic Control
Synthetic Diff. in Differences
160
120
80
40
1970
1980
1990
2000
control
california
40
0
−40
−80
0.05
0.10
0.15
0.20
0.25
Alabama
Arkansas
Colorado
Connecticut
Delaware
Georgia
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Mexico
North Carolina
North Dakota
Ohio
Oklahoma
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
West Virginia
Wisconsin
Wyoming

	DID	SC	SDID
Estimate	-27.4	-19.8	-13.4
Standard error	(16.4)	(7.7)	(7.6)

Table 1: Estimates for average effect of increased cigarette taxes on California per capita cigarette sales over twelve post-treatment years, for difference in differences (DID), synthetic controls (SC), and synthetic difference in differences (SDID), along with an estimated standard error. We discuss the calculation of the standard errors for SDID in Section 5.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American statistical Association* 105 (490):493–505.

Abadie, Alberto and Javier Gardeazabal. 2003. "The economic costs of conflict: A case study of the Basque Country." *American economic review* 93 (1):113–132.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2020. "Synthetic Difference in Differences."

Athey, Susan and Guido W Imbens. 2017. "The state of applied econometrics: Causality and policy evaluation." *Journal of Economic Perspectives* 31 (2):3–32.