Time Series: Synthetic Controls and Difference in Difference

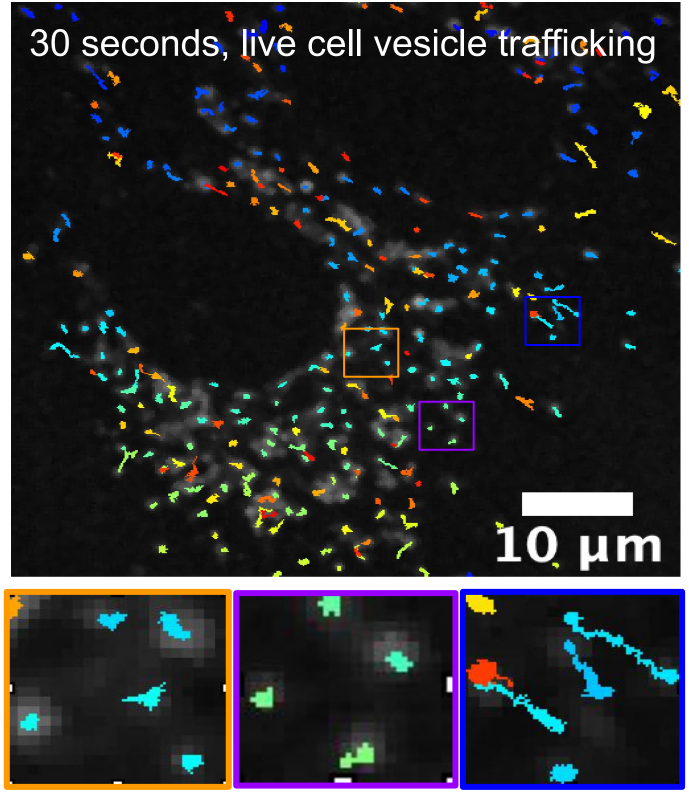
Gaurav Parikh and Nathan Rayens

2/14/23

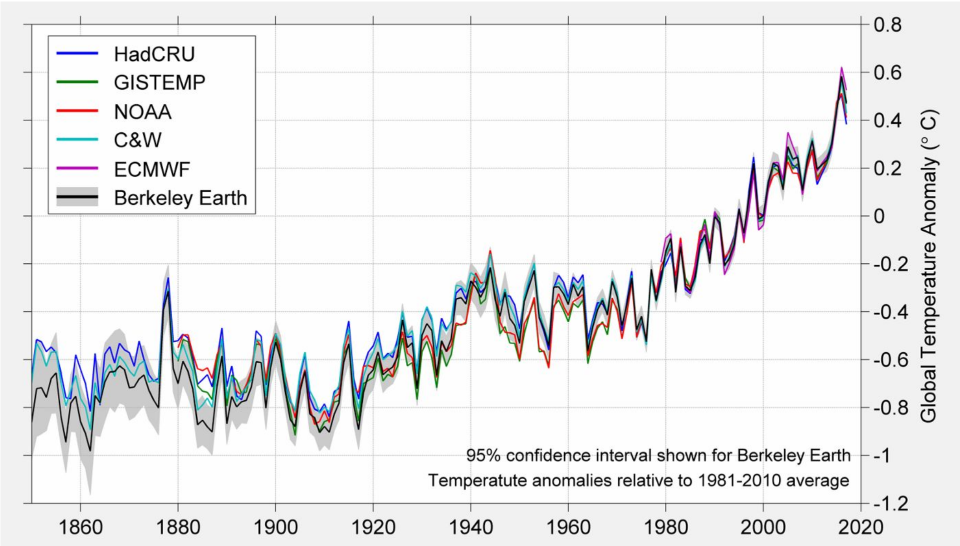
Computer Science 590
Professor Roy

Time Series Data

Tiny time scales



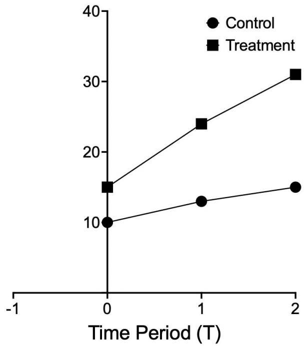
Huge time scales!



Vivek Palaniappan, *Time Series Analysis*

Big Ideas

Difference in Difference



We can't always get a good control in real life. Difference in difference allows ethical and practical comparison of analogous situations to isolate causality of target variables (pollution → cancer, social programs → wellbeing, etc.)

"The difference in difference (DID) design is a quasi-experimental research design that researchers often use to study causal relationships in public health settings where randomized controlled trials (RCTs) are infeasible or unethical. However, causal inference poses many challenges in DID designs... The DID design is not a perfect substitute for randomized experiments, but it often represents a feasible way to learn about casual relationships.

Wing et al., Designing Difference in Difference Studies: Best Practices for Public Health Policy Research, Annual Review of Public Health

Synthetic Controls

"The inclusion of a counterfactual improves causal inference for approaches based on time series analysis, but the selection of a suitable counterfactual or control area can be problematic. The synthetic control method builds a counterfactual using a weighted combination of potential control units"

Bouttell et al., Synthetic control methodology as a tool…, Journal of Epidemiology and Community Health

Synthetic controls in academic research

Table 1 Health-related studies using synthetic control methodology

Date	First author (reference)	Exposure	Outcome	Treated unit(s)	Donor pool
Health finance and health systems reform					
2016	Ryan[4]	Pay for performance	Mortality	UK	Other high-income countries
2015	Mas (unpublished)	Healthcare integration	Healthcare efficiency measures	Spanish integrated healthcare unit	Spanish non-integrated healthcare units
2015	Lepine[5]	User fee removal	Healthcare utilisation	Treated Zambian regions	Untreated Zambian regions
2015	Kreif[6]	Pay for performance	Risk adjusted mortality	Treated UK hospitals	Untreated UK hospitals
2015	Machado[7]	Levels of health insurance	Infertility levels	States with strong infertility mandates	States with weak infertility mandates
2015	Roy[8]	Health reform	Sources of health insurance	Insured Massachusetts Population	Uninsured Massachusetts population
2014	Courtemanche[9]	Levels of health insurance	Self-reported health	Massachusetts	Untreated US states
2014	Dunn[10]	Health reform	Physician payments	Massachusetts	Untreated US states
2014	Tuzemen[11]	Health reform	Non-insurance rate	Massachusetts	Untreated US states
2013	Lo[12]	Income levels	Substitution of public/private insurance	Illinois	Untreated US states
Industry regulation					
2016	Quast[13]	Registration of sex workers	Incidence of sexually transmitted disease	Tijuana, Mexico	Untreated Mexican regions
2015	Restrepo[14]	Trans fat ban	Heart disease mortality	Treated counties in New York	Untreated counties in New York
2014	Sampaio[15]	Mobile phone ban	Road accidents	New York state	Untreated US states
2014	Restrepo[16]	Trans fat ban	Heart disease mortality	Denmark	Other OECD countries
2014	Green[17]	Alcohol licencing hours	Road accidents	England and Wales	Scottish regions
2014	Wang[18]	Pasteurisation of milk	Child mortality	Treated US cities	Untreated US cities
2014	Cunningham[19]	Decriminalisation of indoor prostitution	Incidence of sexually transmitted disease	Rhode Island, USA	Untreated US states/cities

Bouttell et al., Synthetic control methodology as a tool…, Journal of Epidemiology and Community Health

Abadie et al.: Evaluating efficacy of smoking excise tax

Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program

Alberto ABADIE, Alexis DIAMOND, and Jens HAINMUELLER

Building on an idea in Abadie and Gardeazabal (2003), this article investigates the application of synthetic control methods to comparative case studies. We discuss the advantages of these methods and apply them to study the effects of Proposition 99, a large-scale tobacco control program that California implemented in 1988. We demonstrate that, following Proposition 99, tobacco consumption fell markedly in California relative to a comparable synthetic control region. We estimate that by the year 2000 annual per-capita cigarette sales in California were about 26 packs lower than what they would have been in the absence of Proposition 99. Using new inferential methods proposed in this article, we demonstrate the significance of our estimates. Given that many policy interventions and events of interest in social sciences take place at an aggregate level (countries, regions, cities, etc.) and affect a small number of aggregate units, the potential applicability of synthetic control methods to comparative case studies is very large, especially in situations where traditional regression methods are not appropriate.

KEY WORDS: Observational studies; Proposition 99; Tobacco control legislation; Treatment effects.

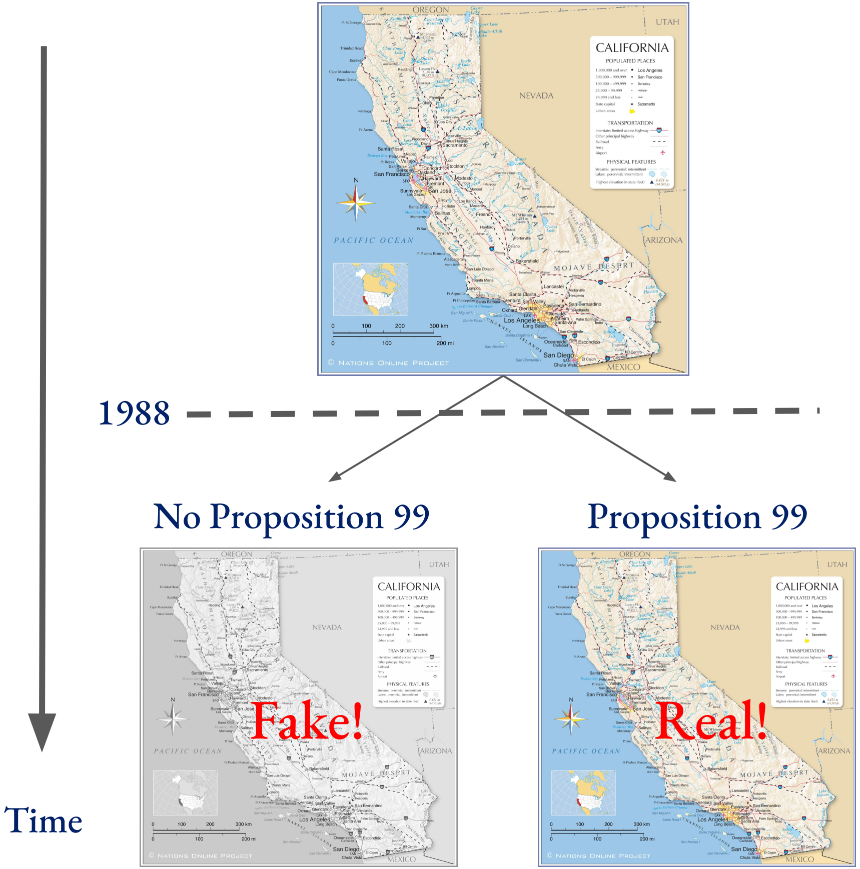


California Proposition 99		
Result	Votes	Percentage
✔ Yes	**5,607,387**	58.17%
No	4,032,644	41.83%

Proposition 99, 1988, Ballotpedia

Into the California-verse: challenges of time series data

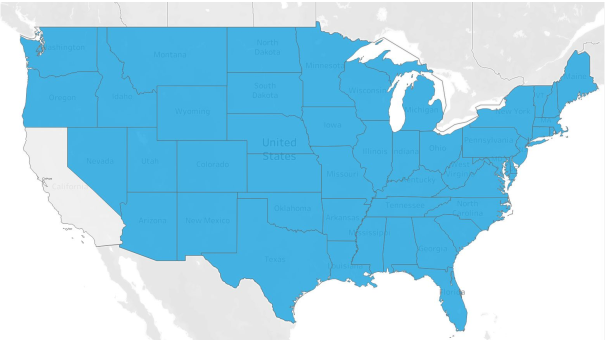
- Sometimes, good controls don't exist!
  - **What is the alternative to real California?**
- Turns out, synthetic control generation can save the day:
  - Transparent and easy to interpret
  - No need for post-intervention data
- **Priority is making a synthetic control that mimics the real data up until intervention**



How can we make this synthetic control?

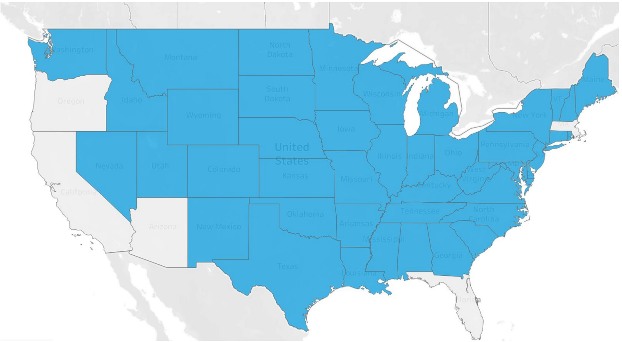
Building a control pool for weighted synthesis

Equally weight **real** post-intervention data from every other state



J = 50
(49 + DC)

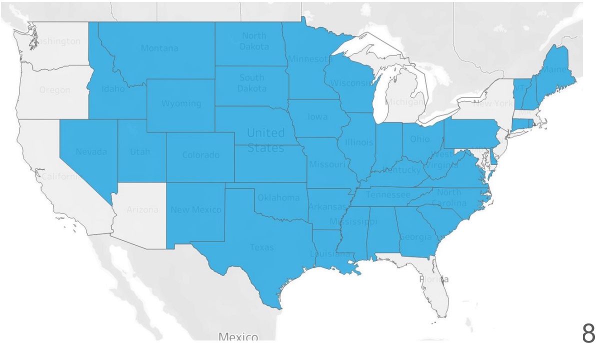
Remove states that enacted Prop 99 analogs in the timeframe



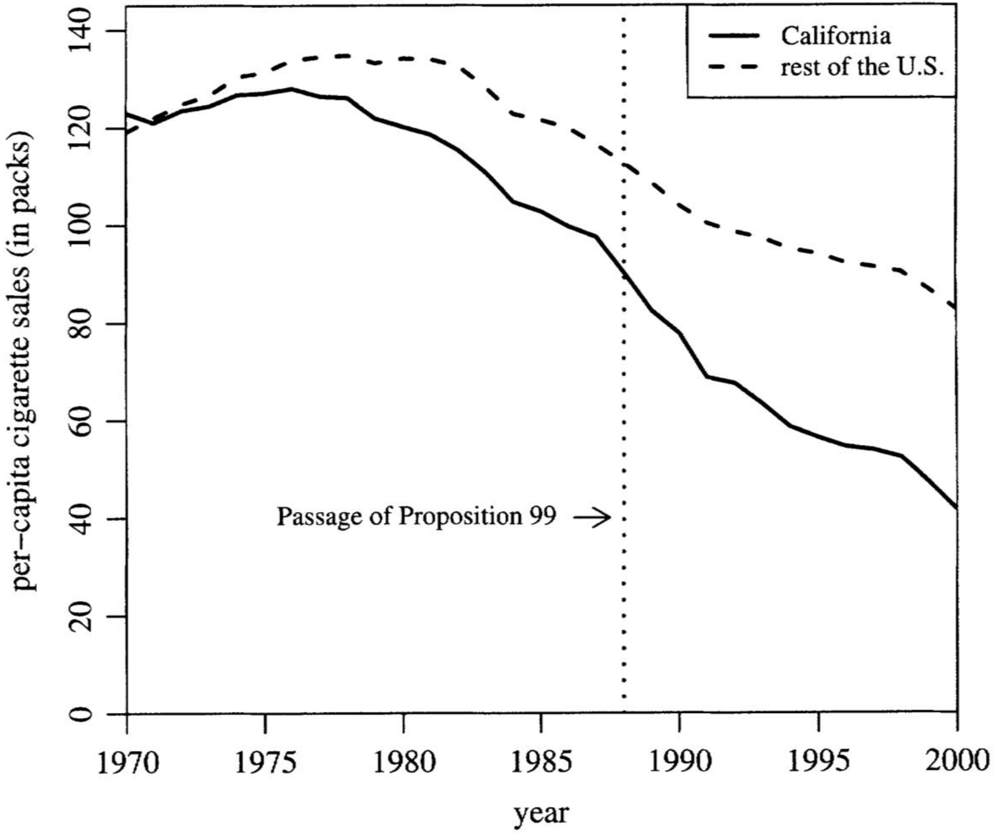
J = 46
(45 + DC)

J = 38

Remove states that increased cigarette taxes by > $0.50

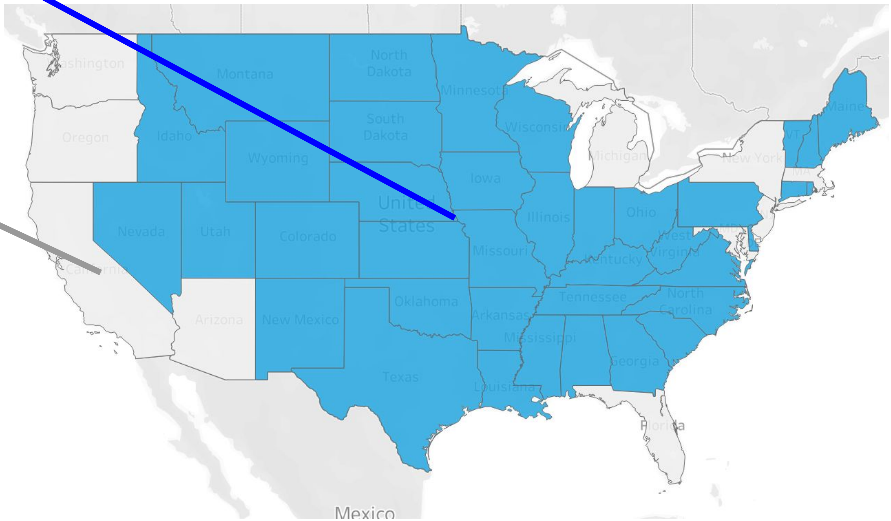


Even still, the US is not a good control for California!



Introducing modeling terminology

Y_{it}^{N}	Outcome of unaffected state i at time t, ***without intervention***
$1 \leq T_0 < T$	T_0 is when intervention occurs, T is end of data
Y_{it}^{I}	Outcome of affected state i at time t
$Y_{it}^{N} = Y_{it}^{I}$	$t \in \{1 \ldots T_0\} i \in \{1 \ldots N\}$
$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$	Intervention effect at time t in state i
$\alpha_{1t} = Y_{1t} - Y_{1t}^{N}$	Intervention effect after time T_0 with one true affected state



T_0 = 1988, so we assume no difference between California and the rest of US prior to Prop 99 intervention

Under the hood (supported by Appendix B)

Proposed factor model for
estimate unobserved outcomes

[1 x r] vector of unknown parameters

Unobserved common factors

$$Y_{it}^N = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_i + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \varepsilon_{it}$$

Observed [r x 1] vector of covariates

Unobserved shocks

Shared, unknown common factor loading

Unknown common factor loading

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \delta_t + \boldsymbol{\theta}_t \sum_{j=2}^{J+1} w_j \mathbf{Z}_j + \boldsymbol{\lambda}_t \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}.$$

Apply a weight vector, W = {w_2, … w_{J+1}}' affecting all of the non-intervention states with weights summing to 1

Now, find good weights, W*, that approximate pre-intervention California

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}, \quad \sum_{j=2}^{J+1} w_j^* Y_{j2} = Y_{12}, \quad \ldots, \quad \sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0}, \quad \text{and} \quad \sum_{j=2}^{J+1} w_j^* \mathbf{Z}_j = \mathbf{Z}_1.$$

In Appendix B, the authors derive the following expansion of these terms:

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt} = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \boldsymbol{\lambda}_t \left(\sum_{n=1}^{T_0} \boldsymbol{\lambda}_n' \boldsymbol{\lambda}_n \right)^{-1} \boldsymbol{\lambda}_s' (\varepsilon_{js} - \varepsilon_{1s}) - \sum_{j=2}^{J+1} w_j^* (\varepsilon_{jt} - \varepsilon_{1t}).$$

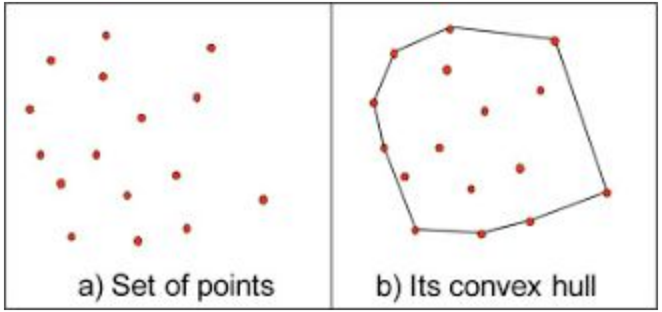
With a significant amount of pain, the authors then show that we end up sending our bias for the synthetic control comparison to 0 as we increase the number of periods pre-intervention, leaving the following:

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

Some caveats to this relationship

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

The authors note that this only works if Z is accurately mapped by the synthetic control and that there may be cases where the synthetic control wouldn't resolve this problem, such as if the treatment region lies outside the convex hull of the donors



Laurini, *Geographic Knowledge Infrastructure*

This can be checked case-by-case *in situ* so the authors recommend performing that check prior to going further with synthetic controls

$$\sum_{j=2}^{J+1} w_j^* \mathbf{Z}_j = \mathbf{Z}_1 \quad \text{and} \quad \sum_{j=2}^{J+1} w_j^* \boldsymbol{\mu}_j = \boldsymbol{\mu}_1,$$

- We can't see the terms encapsulated by μ
- Still, fitting well to state covariates, Z, implies good fitting of μ

That painful work from Appendix B

Here, the authors bound the factor equation to pre-intervention period, P, and then bound the unobserved common factors within this pre-intervention time period

Original expression

$$Y_{it}^N = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_i + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \varepsilon_{it}$$

Subtracting from the initial equation, we get the following:

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j Y_{jt}^N = \boldsymbol{\theta}_t \left(\mathbf{Z}_1 - \sum_{j=2}^{J+1} w_j \mathbf{Z}_j \right) + \boldsymbol{\lambda}_t \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j \right) + \sum_{j=2}^{J+1} w_j (\varepsilon_{1t} - \varepsilon_{jt}).$$

Isolate to T_0 length vectors for pre-intervention, P

$$\mathbf{Y}_1^P - \sum_{j=2}^{J+1} w_j \mathbf{Y}_j^P = \boldsymbol{\theta}^P \left(\mathbf{Z}_1 - \sum_{j=2}^{J+1} w_j \mathbf{Z}_j \right) + \boldsymbol{\lambda}^P \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} w_j \boldsymbol{\mu}_j \right) + \sum_{j=2}^{J+1} w_j (\boldsymbol{\varepsilon}_1^P - \boldsymbol{\varepsilon}_j^P).$$

$$\frac{1}{M} \sum_{t=T_0-M+1}^{T_0} \boldsymbol{\lambda}_t' \boldsymbol{\lambda}_t.$$

Assume all unobserved common factors will be less than or equal to some unknown common factor lambda bar

$$|\lambda_{tf}| \leq \bar{\lambda} \text{ for all } t = 1, \ldots, T, f = 1, \ldots, F.$$

That painful work from Appendix B

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}, \quad \sum_{j=2}^{J+1} w_j^* Y_{j2} = Y_{12}, \quad \ldots, \quad \sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0}, \quad \text{and} \quad \sum_{j=2}^{J+1} w_j^* \mathbf{Z}_j = \mathbf{Z}_1.$$

Then, the authors show that with appropriate weights, w* and beyond T_0, some of their bias terms go away or are centered at 0 mean.

Simplify notation as annotated w/ R_{2t} and R_{3t} centered at mean 0 beyond intervention. ***Note the Y and Z terms go away by nature of our w* requirement***

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N = R_{1t} + \cancelto{0}{R_{2t}} + \cancelto{0}{R_{3t}},$$

Apply this relationship to the original equation to generalize relationship

$$\begin{aligned}
Y_{1t}^N - \sum_{j=2}^{J+1} w_j Y_{jt}^N &= \boldsymbol{\lambda}_t (\boldsymbol{\lambda}^{P\prime} \boldsymbol{\lambda}^P)^{-1} \boldsymbol{\lambda}^{P\prime} \left(\mathbf{Y}_1^P - \sum_{j=2}^{J+1} w_j \mathbf{Y}_j^P \right) \\
&\quad + (\boldsymbol{\theta}_t - \boldsymbol{\lambda}_t (\boldsymbol{\lambda}^{P\prime} \boldsymbol{\lambda}^P)^{-1} \boldsymbol{\lambda}^{P\prime} \boldsymbol{\theta}^P) \left(\mathbf{Z}_1 - \sum_{j=2}^{J+1} w_j \mathbf{Z}_j \right) \\
&\quad - \boldsymbol{\lambda}_t (\boldsymbol{\lambda}^{P\prime} \boldsymbol{\lambda}^P)^{-1} \boldsymbol{\lambda}^{P\prime} \left(\boldsymbol{\varepsilon}_1^P - \sum_{j=2}^{J+1} w_j \boldsymbol{\varepsilon}_j^P \right) \\
&\quad + \sum_{j=2}^{J+1} w_j (\varepsilon_{1t} - \varepsilon_{jt}).
\end{aligned}$$

R_{2t} R_{1t}

R_{3t}

Expand the remaining term back out for all t

$$R_{1t} = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \boldsymbol{\lambda}_t \left(\sum_{n=1}^{T_0} \boldsymbol{\lambda}_n' \boldsymbol{\lambda}_n \right)^{-1} \boldsymbol{\lambda}_s' \varepsilon_{js}.$$

And set the rightmost terms like so

$$\bar{\varepsilon}_j^L = \sum_{s=1}^{T_0} \boldsymbol{\lambda}_t \left(\sum_{n=1}^{T_0} \boldsymbol{\lambda}_n' \boldsymbol{\lambda}_n \right)^{-1} \boldsymbol{\lambda}_s' \varepsilon_{js}$$

That painful work from Appendix B

Jumping back in, we apply Hölder's inequality (for positive, even moments, p) once:

$$\sum_{j=2}^{J+1} w_j^* |\bar{\varepsilon}_j^L| \leq \left(\sum_{j=2}^{J+1} w_j^* |\bar{\varepsilon}_j^L|^p\right)^{1/p} \leq \left(\sum_{j=2}^{J+1} |\bar{\varepsilon}_j^L|^p\right)^{1/p}.$$

And then again:

$$E\left[\sum_{j=2}^{J+1} w_j^* |\bar{\varepsilon}_j^L|\right] \leq \left(E\left[\sum_{j=2}^{J+1} |\bar{\varepsilon}_j^L|^p\right]\right)^{1/p}.$$

And then Rosenthal's inequality:

$$E|\bar{\varepsilon}_j^L|^p \leq C(p)\left(\frac{\bar{\lambda}^2 F}{\underline{\xi}}\right)^p \max\left\{\frac{1}{T_0^p}\sum_{t=1}^{T_0} E|\varepsilon_{jt}|^p, \left(\frac{1}{T_0^2}\sum_{t=1}^{T_0} E|\varepsilon_{jt}|^2\right)^{p/2}\right\},$$

This is goes to 0 for large pre-intervention T

- Authors apply Hölder's inequality twice
- Then Rosenthal's inequality
- Successfully bound the expected bias of the predictive factor model at 0 for large pre-intervention time periods
    - Assumes relatively small independent state shocks.

All that to say → we can use a simple linear combination of observed factors to make a low-bias estimate of California

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

Implementation (finally!)

$$\bar{Y}_i^{\mathbf{K}} = \sum_{s=1}^{T_0} k_s Y_{is}$$

Time-scaled factor of interest, Y, weighting by k periods of exposure

$$\mathbf{K}_1, \dots, \mathbf{K}_M$$

Encapsulated in M linear combinations to represent target variable linear combinations

$\bar{Y}_i^{\mathbf{K}_1}, \dots, \bar{Y}_i^{\mathbf{K}_M}$ is $\bar{Y}_i^{\mathbf{K}_1} = Y_{i1}, \dots, \bar{Y}_i^{\mathbf{K}_{T_0}} = Y_{iT_0}$

Ln(GDP per capita)
Percent aged 15–24
Retail price
Beer consumption per capita

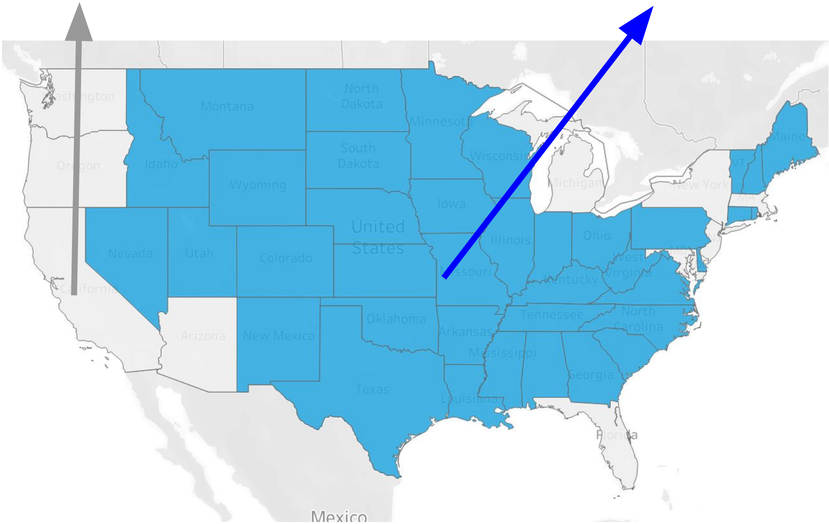
$$\mathbf{X}_1 = (\mathbf{Z}_1', \bar{Y}_1^{\mathbf{K}_1}, \dots, \bar{Y}_1^{\mathbf{K}_M})'$$

Set of pre-intervention parameters from California

$\mathbf{X}_0$ is a $(k \times J)$ matrix

$$(\mathbf{Z}_j', \bar{Y}_j^{\mathbf{K}_1}, \dots, \bar{Y}_j^{\mathbf{K}_M})'$$

Set of J control states' pre-intervention parameters



Implementation

$$\mathbf{X}_1 = (\mathbf{Z}_1', \bar{Y}_1^{\mathbf{K}_1}, \ldots, \bar{Y}_1^{\mathbf{K}_M})'$$

Set of pre-intervention parameters from California

$\mathbf{X}_0$ is a $(k \times J)$ matrix

$$(\mathbf{Z}_j', \bar{Y}_j^{\mathbf{K}_1}, \ldots, \bar{Y}_j^{\mathbf{K}_M})'$$

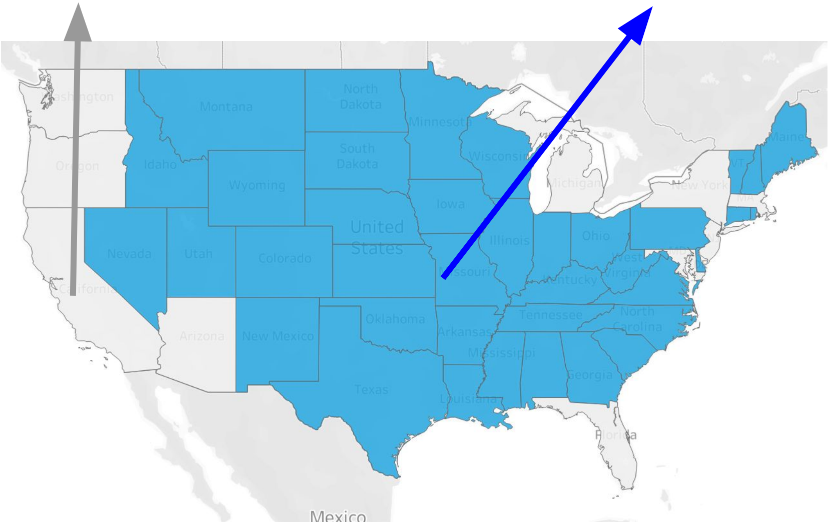
Set of J control states' pre-intervention parameters

$$\mathbf{W} = (w_2, \ldots, w_{J+1})'$$

$$w_j \geq 0 \text{ for } j = 2, \ldots, J+1$$

$$w_2 + \cdots + w_{J+1} = 1$$

Weights for the states of interest



Minimize this!

$$\|\mathbf{X}_1 - \mathbf{X}_0\mathbf{W}\|$$

Four core predictors for synthetic California

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

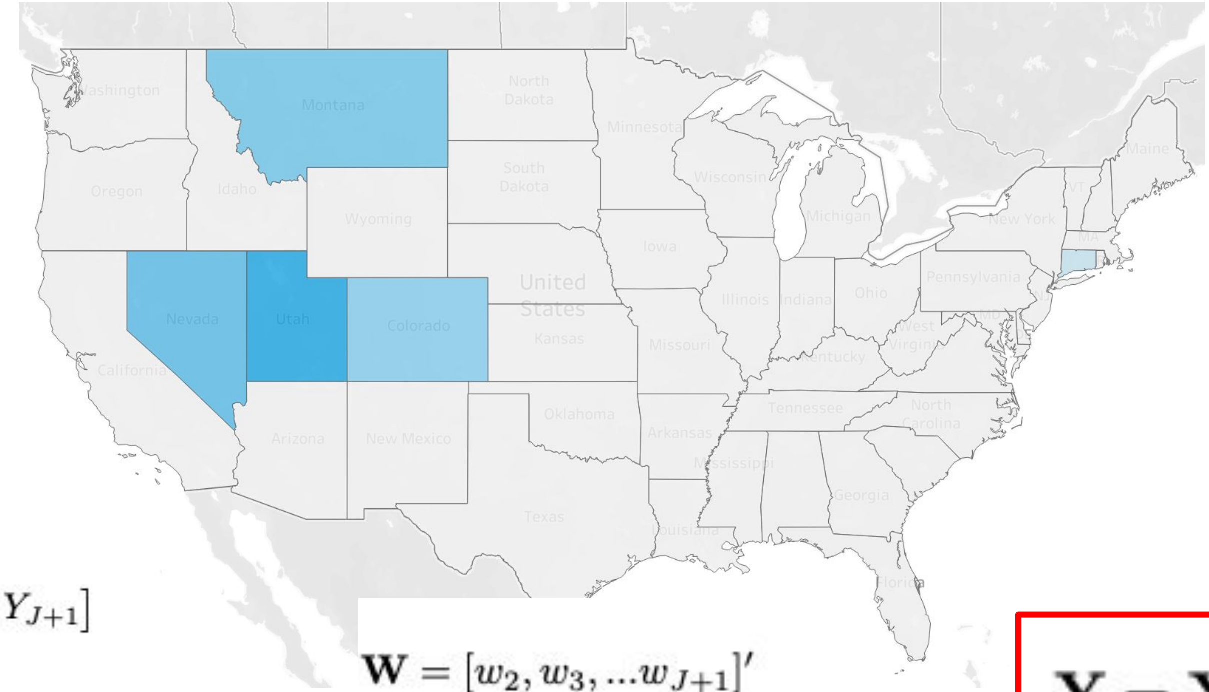
Now that we have a synthetic control set, how can we predict cigarette sales?

(For those who skimmed the paper) Any guesses on what states make a synthetic California?

Synthetic California = West Coast and … Connecticut

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



$$\mathbf{X}_1 = \begin{bmatrix} Y_2 & \ldots & Y_{J+1} \end{bmatrix}$$

Set of J control states' cigarette sales per capita

Fixed and real data

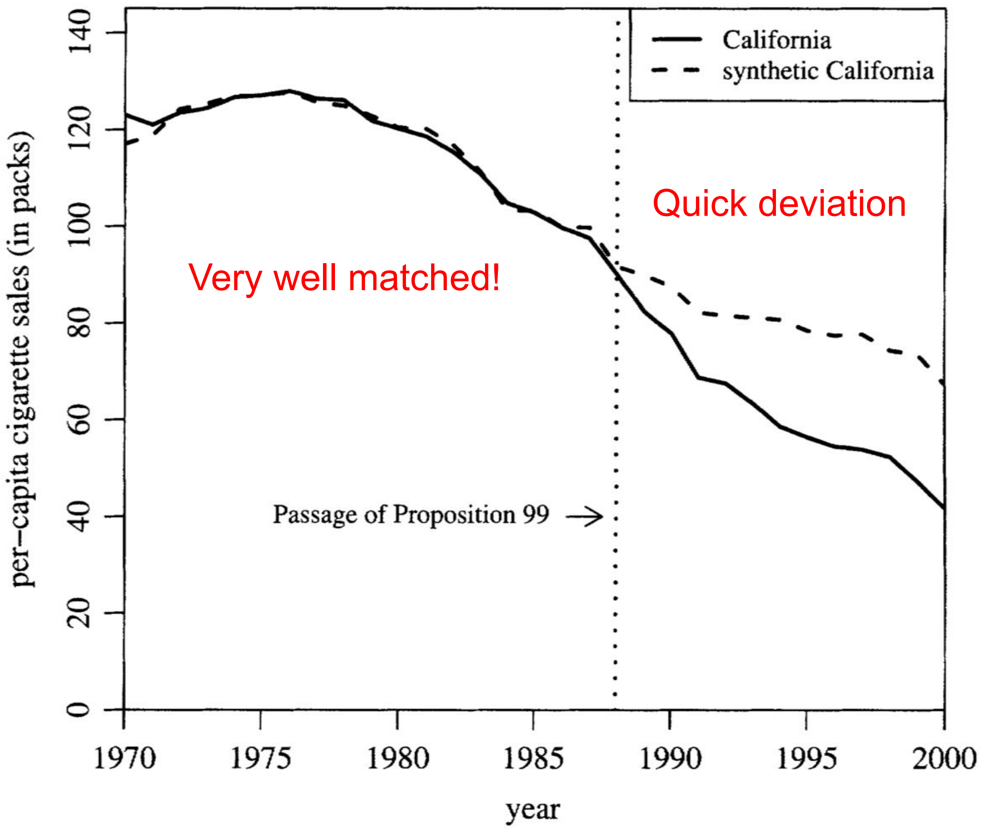
$$\mathbf{W} = [w_2, w_3, \ldots w_{J+1}]'$$

Set of weights for J control states

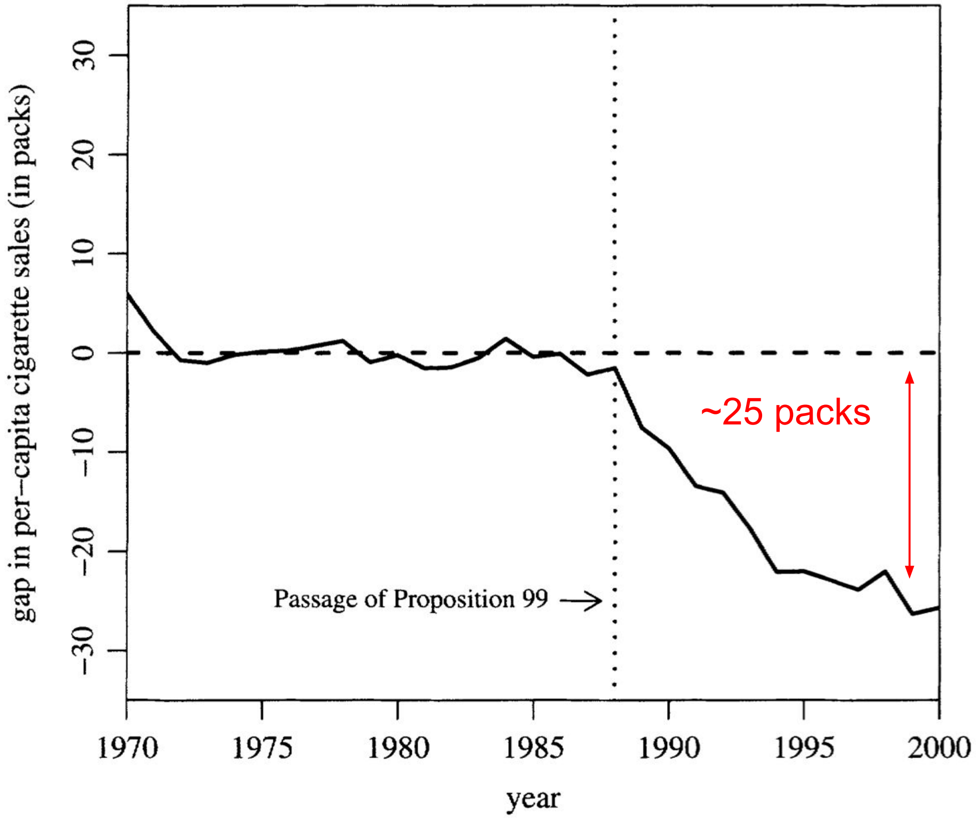
Now fixed!

$$\boxed{\mathbf{Y} = \mathbf{X}_1\mathbf{W}}$$

Synthetic California lags real California in cigarette decline



Prop 99 decreased cigarette sales by ~25 packs per capita

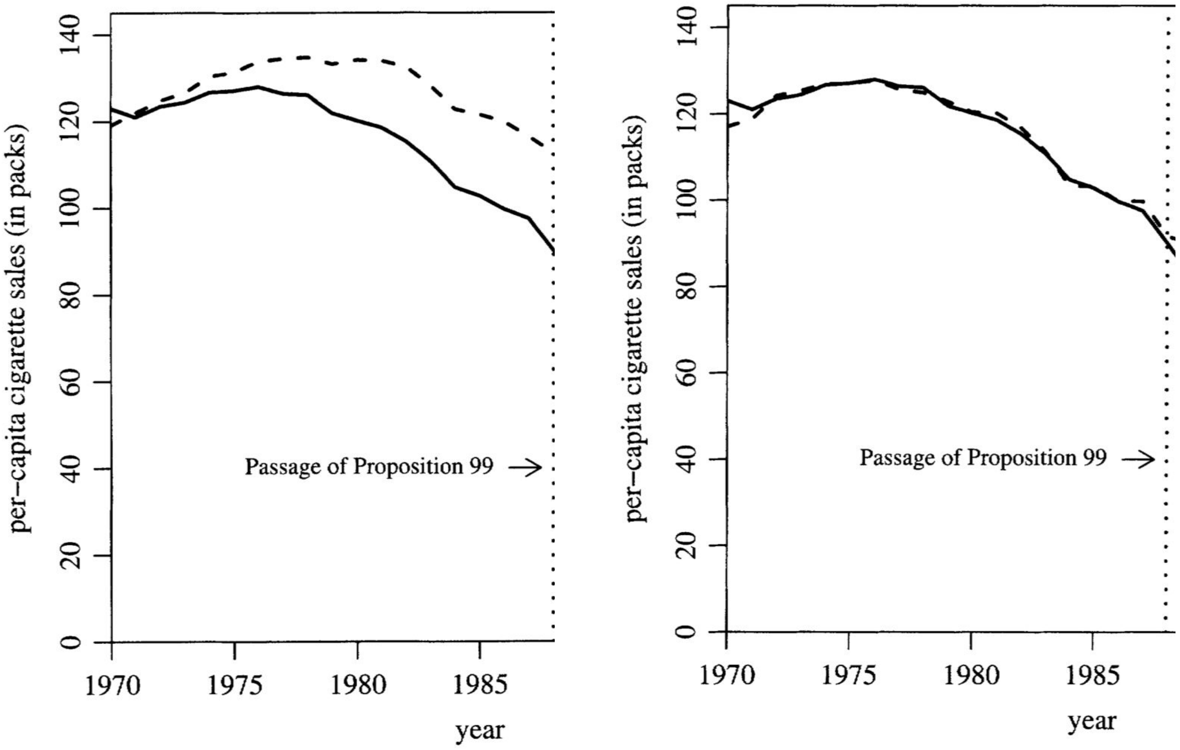


A quick note on mean squared prediction error (MSPE)

The authors use MSPE over the ***pre-treatment*** time periods to check that the modeling is informative

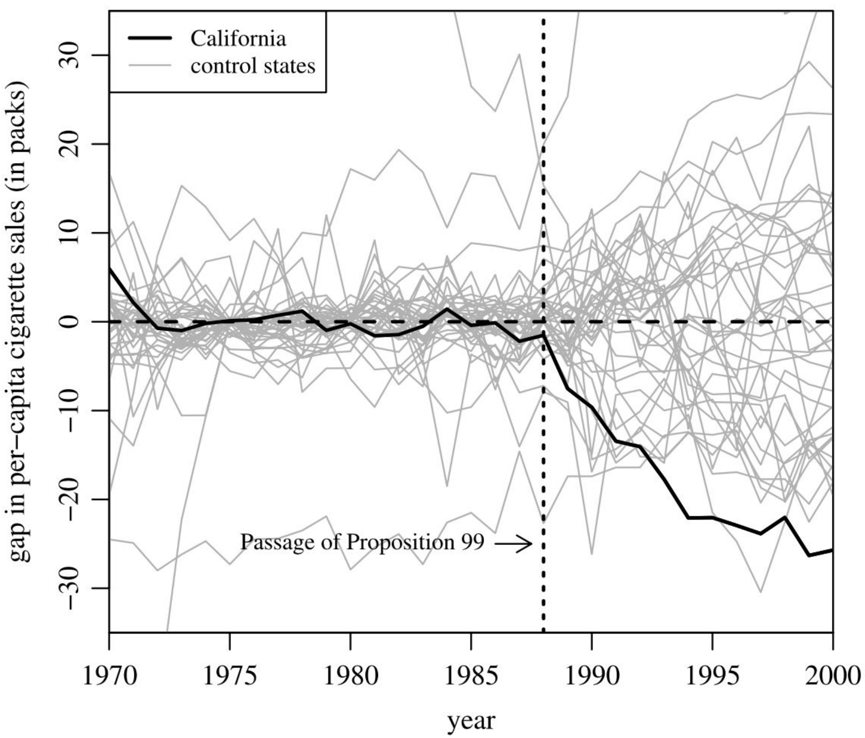
$$MSPE = \frac{1}{T_0} \sum_{t=1}^{T_0} (Y_{true,t} - Y_{predicted,t})^2$$

High MSPE means we are not mapping to the pretreatment region successfully!

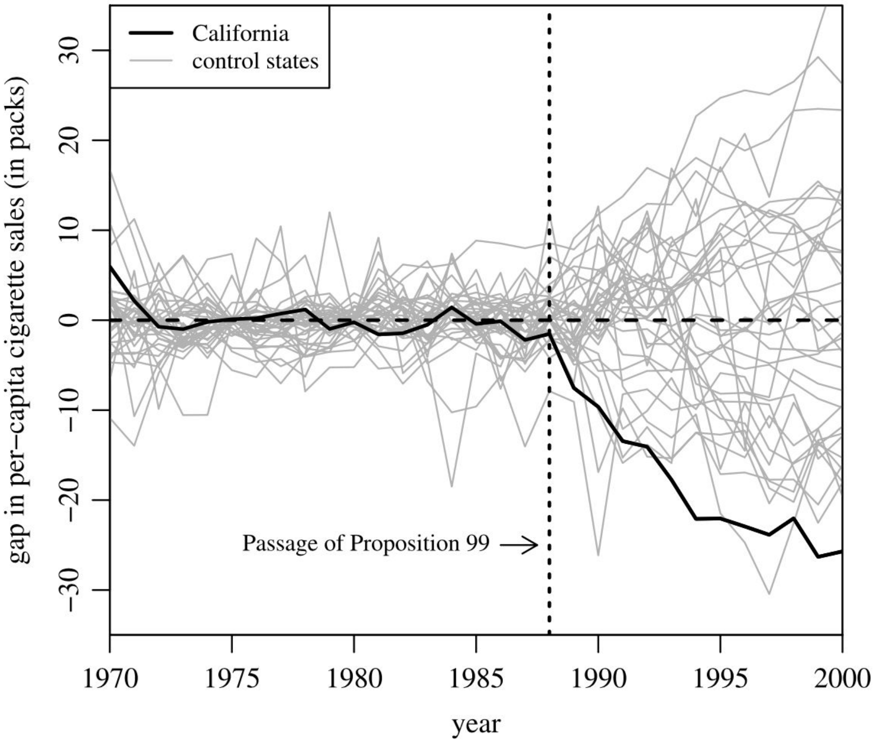


How do we know the method is generalizable and not California-specific?

Placebo controls show metric is strong, not infallible



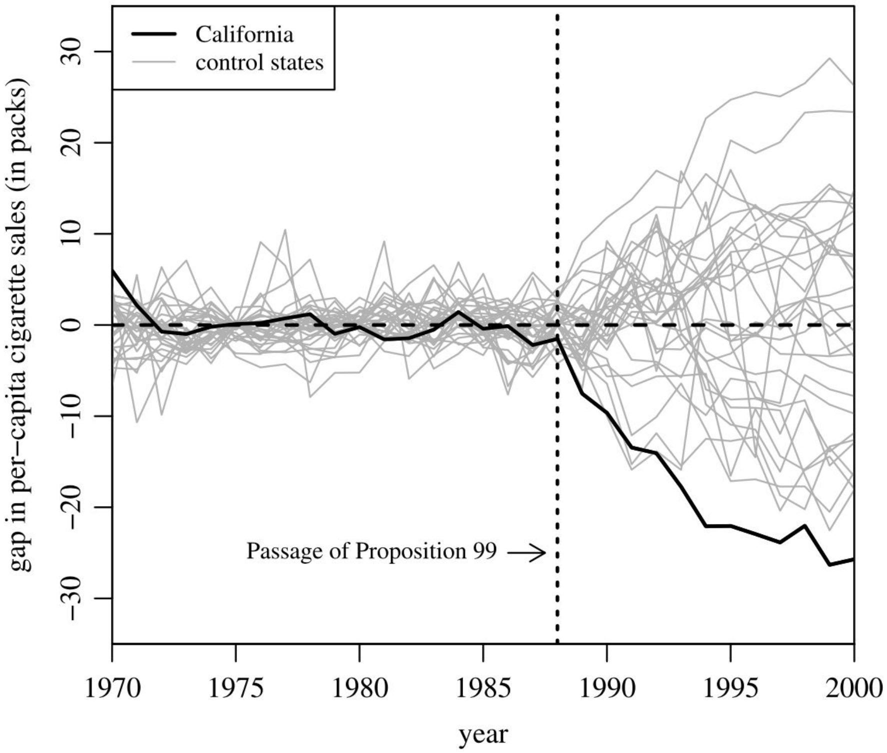
Raw data



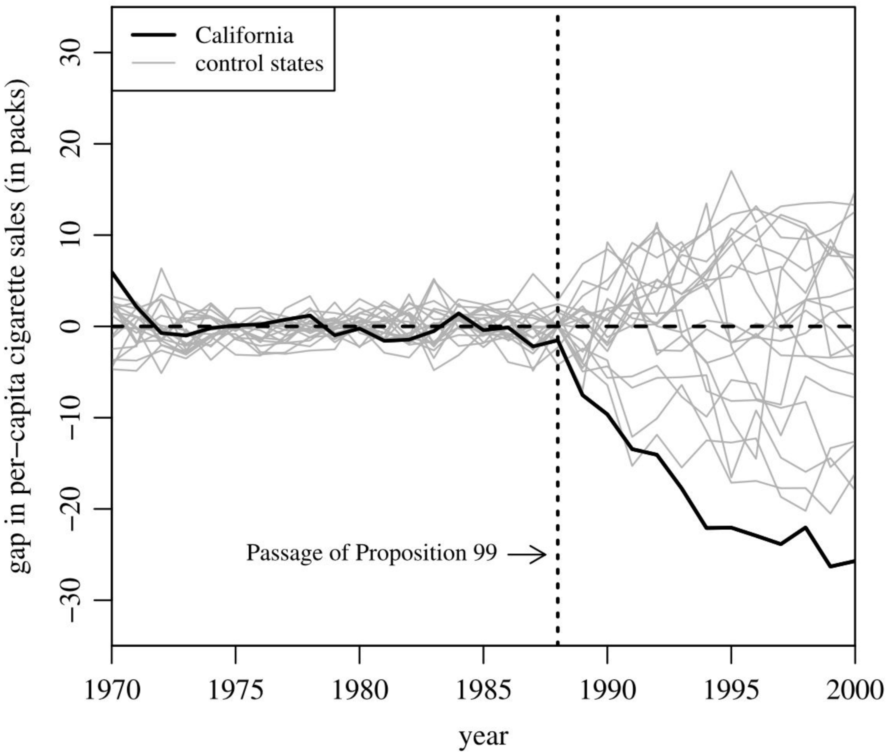
Filtering placebo MSPE > 20x California

Any guesses on the worst performing state for this placebo control method?

Placebo controls show metric is strong, not infallible

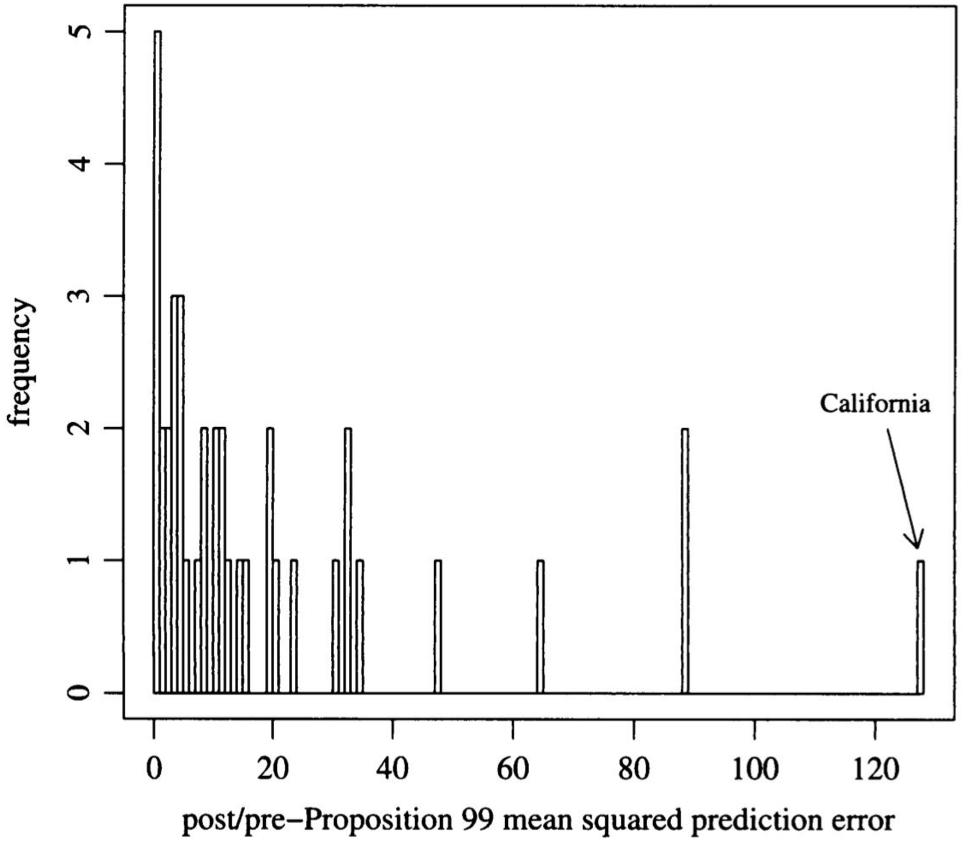


Filtering placebo MSPE > 5x California



Filtering placebo MSPE > 2x California

California separates in MSPE pre/post intervention ratio



What did we learn?

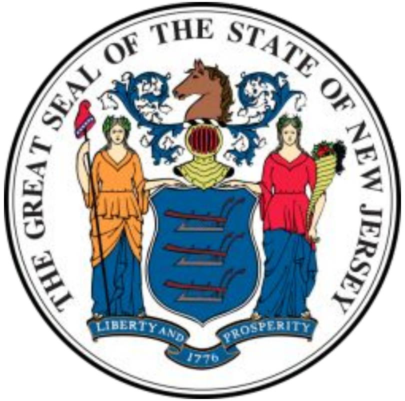
- For successful comparisons, you need a good control set
- Synthetic controls can be generated from untreated analogs
- **Simple linear combination works!**
- Expert knowledge and/or wide variable set may be required to tailor control pool to the target variable
- **Covering time variant bases can be extremely labor intensive**
- Placebo controls for control pool can help validate the synthesis method

Card and Krueger. Determining if increasing minimum wage leads to increase in unemployment

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

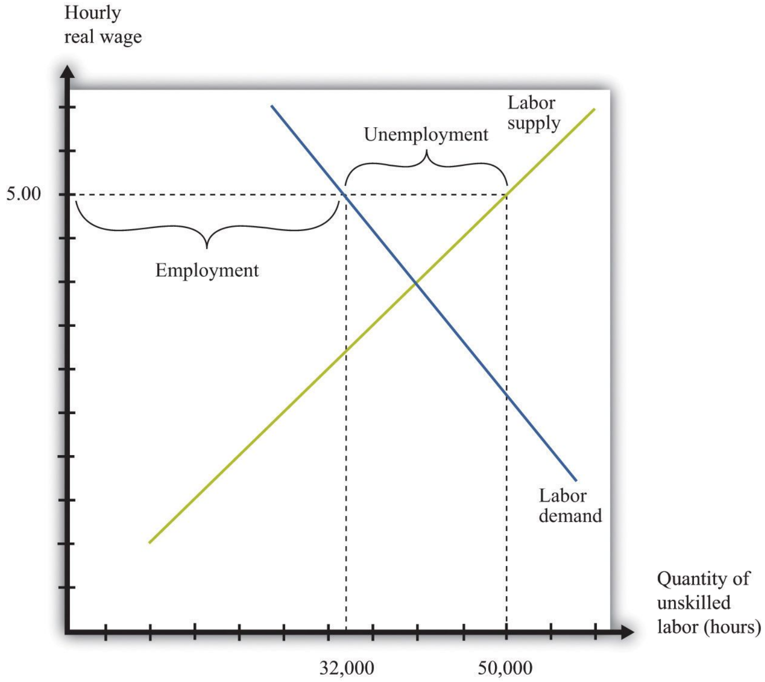
By DAVID CARD AND ALAN B. KRUEGER*

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)



Traditional Economic View and Motivation

- Economists believed
  - **Increase in wage means that employers must pay more**
  - **Employers tend to hire less in face of rising cost of labour**
- **As a result, expectation of decline in full time workers.**
- How to measure if this is true, in the midst of trends like
  - **Recessionary economy**



Card and Kreuger: Timeline and Study Design

WAVE 1: Survey of 473 stores in NJ(364), PA(109)

FEB-MAR , 1992

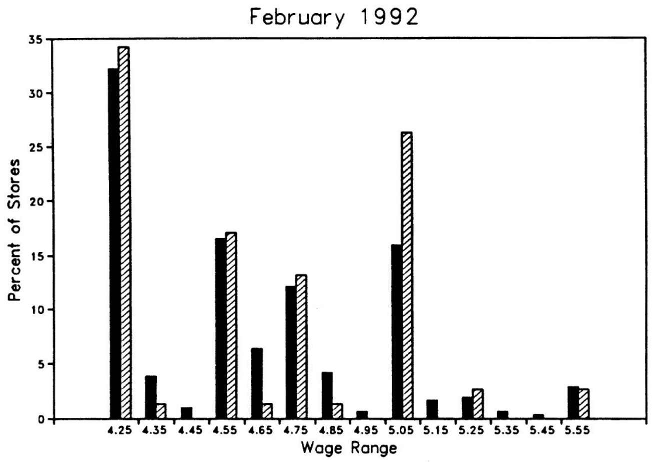
MIN WAGE CHANGED FROM

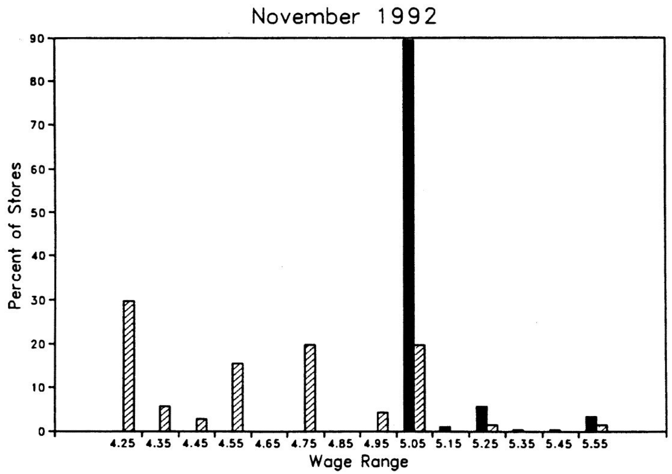
$4.25 to $ 5.25

APR, 1992

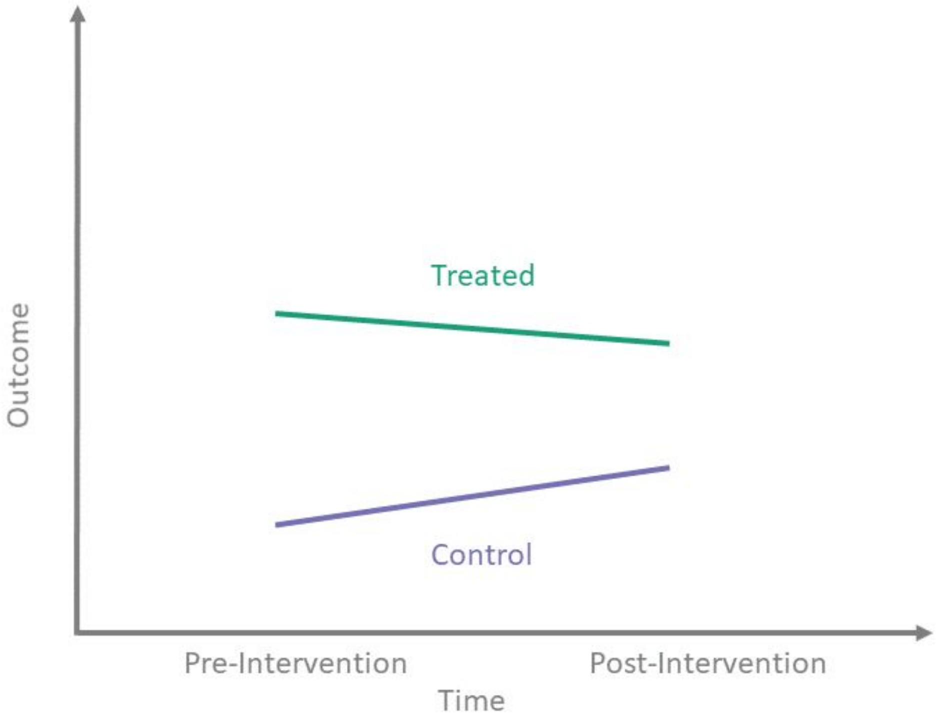
WAVE 2: Called the same 473 stores
Response from 410 stores in
NJ(331), PA(79)

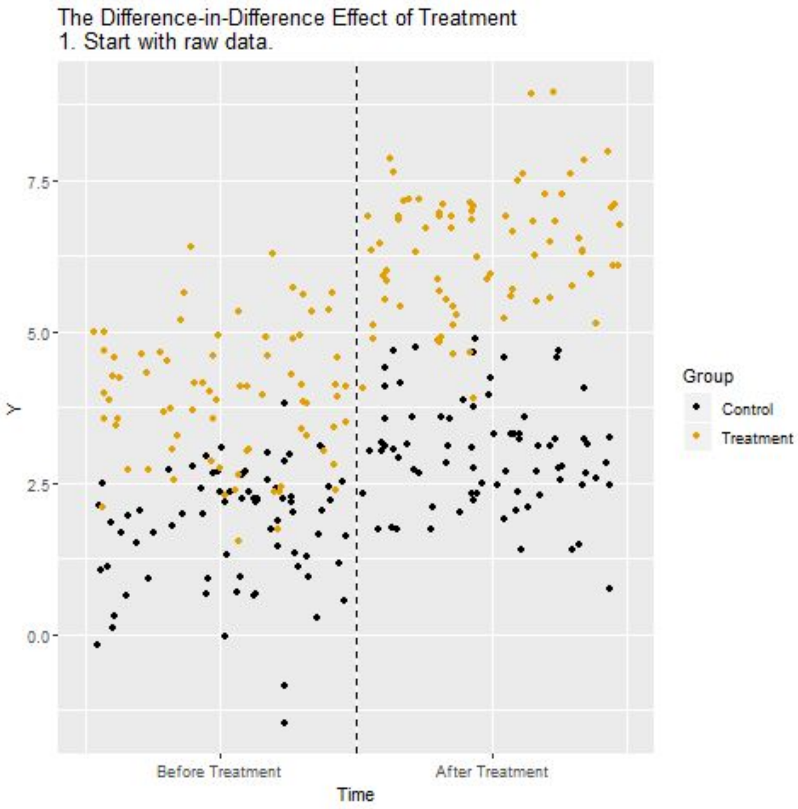
NOV-DEC,1992





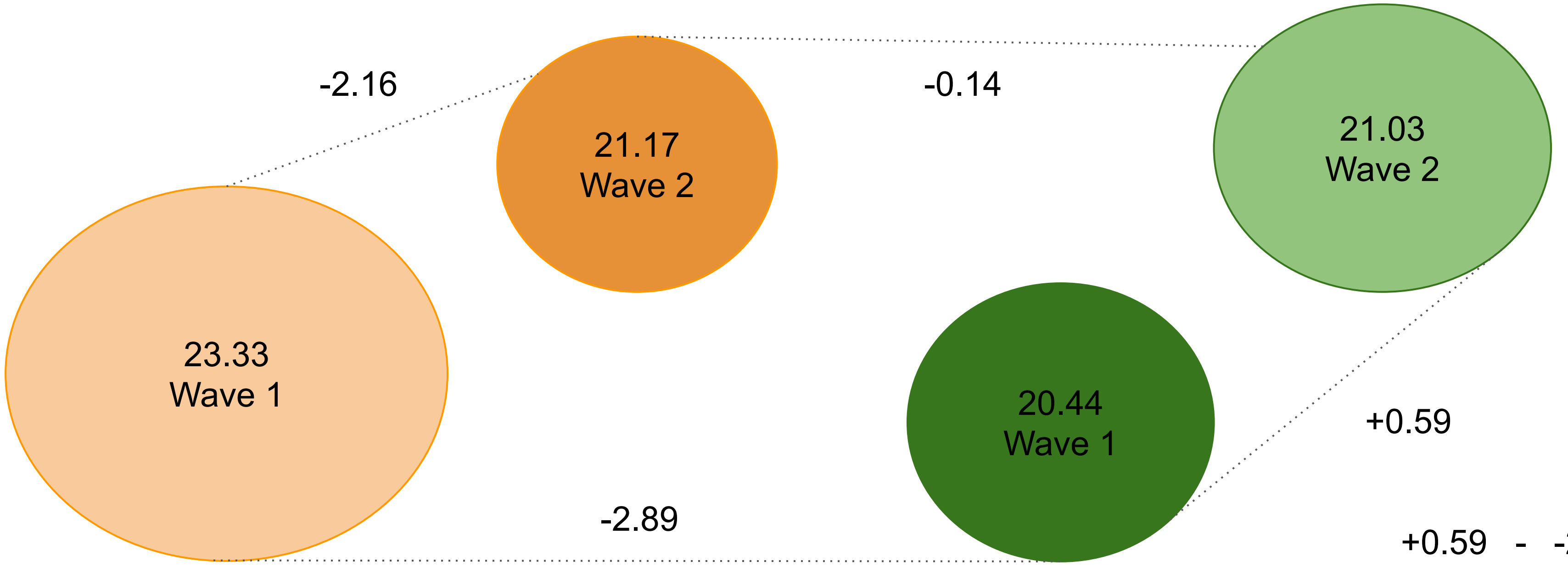
Card and Kreuger: Difference in Differences





Card and Kreuger: Treatment and Control

HOW TO FIND EFFECT OF MINIMUM WAGE CHANGE IN NJ?



EASTERN PENNSYLVANIA

NEW JERSEY

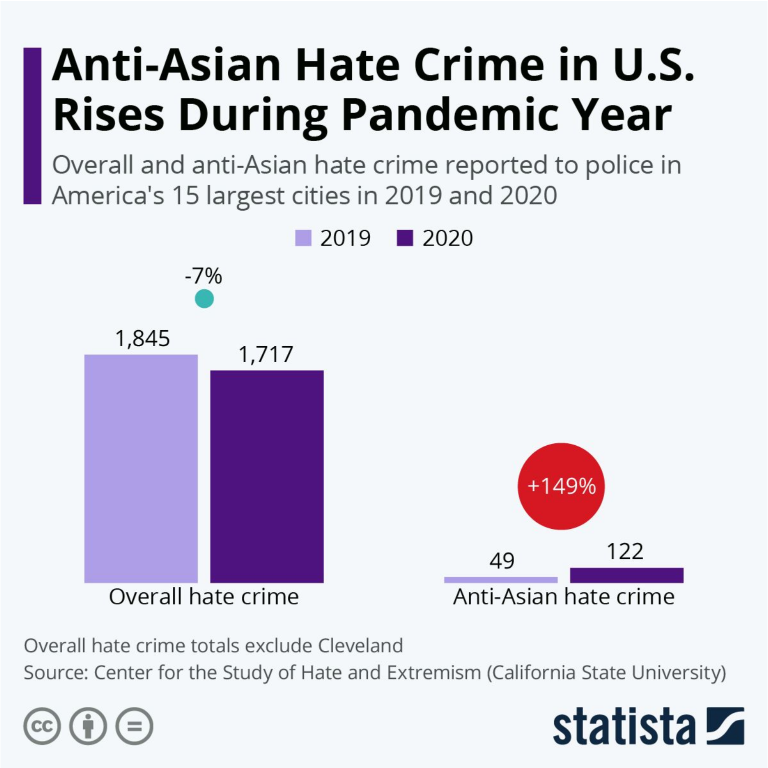
Average Full Time Employees

+0.59 - -2.16

-0.14 - -2.89

2.76

In Class Exercise



Card and Kreuger: Regressing for other factors

$$\Delta E_i = a + b\,\mathbf{X}_i + c\,NJ_i + \varepsilon$$

1.Chain or not
2. Control for region

$$\Delta Ei = a + b\mathbf{X}_i + c\,GAP_i + \varepsilon$$

Wage Gap before increase in min wage

$$GAP_i = \begin{cases} \mathbf{0} \ \ \textbf{for stores in PA} \\ \mathbf{0} \ \ \textbf{for stores in NJ where } \mathbf{W_{1i} > 5.05} \\ \mathbf{(5.05 - W_{1i})/W_{1i}} \ \ \textbf{otherwise} \end{cases}$$

Card and Kreuger: Other aspects of the paper

- **Nonwage offsets:**

  higher min wage may lead to reduced employee benefits

  increase in "full meal" food prices

  no significant effects on store openings

Difference in Differences: Potential Outcomes Framework

<table>
<tr><th></th><th></th><th colspan="4">Potential outcomes</th></tr>
<tr><th></th><th></th><th colspan="2">Pre-intervention</th><th colspan="2">Post-intervention</th></tr>
<tr><th>Individual</th><th>Group</th><th>Untreated</th><th>Treated</th><th>Untreated</th><th>Treated</th></tr>
<tr><td>1</td><td>Treated</td><td colspan="2">✓</td><td>?</td><td>✓</td></tr>
<tr><td>2</td><td>Treated</td><td colspan="2">✓</td><td>?</td><td>✓</td></tr>
<tr><td>3</td><td>Treated</td><td colspan="2">✓</td><td>?</td><td>✓</td></tr>
<tr><td>…</td><td></td><td colspan="2"></td><td></td><td></td></tr>
<tr><td>N-2</td><td>Control</td><td colspan="2">✓</td><td>✓</td><td>?</td></tr>
<tr><td>N-1</td><td>Control</td><td colspan="2">✓</td><td>✓</td><td>?</td></tr>
<tr><td>N</td><td>Control</td><td colspan="2">✓</td><td>✓</td><td>?</td></tr>
</table>

Notation

Symbol	Meaning
$Y(t)$	Observed outcome at time t
$A = 0$	Control
$A = 1$	Treated
$t = 1, \ldots, T_0$	Pre-treatment times
$t = T_0 + 1, \ldots, T$	Post-treatment times
$Y^a(t)$	Potential outcome with treatment $A = a$ at time t
X	Observed covariates
U	Unobserved covariates

Assumptions

Consistency Assumption

$Y(t)=(1-A)\cdot Y_0(t)+A\cdot Y_1(t)$

If a unit is treated $(A=1)$, then the observed outcome is the potential outcome with treatment $Y(t)=Y_1(t)$ and the potential outcome with no treatment $Y_0(t)$ is unobserved. If a unit is not treated $(A=0)$, then $Y(t)=Y_0(t)$ and $Y_1(t)$ is unobserved.

Time assumption

$Y(t)=Y_0(t)=Y_1(t)$, for $t \leq T_0$,

Before treatment, control and treatment are the same

Counterfactual Assumption

$E[Y_0(2)-Y_0(1) \mid A=1] = E[Y_0(2)-Y_0(1) \mid A=0]$

The change in outcomes from pre- to post-intervention in the control group is a good proxy for the *counterfactual* change in untreated potential outcomes in the treated group

What did we learn?

- **For successful comparisons, you need a good control set**
- Difference in Differences assumes that a control can be found by looking at relative changes between two units where one received treatment
- **Parallel trend assumption must hold**
- Difference in Differences is very useful in settings where it might be hard to randomize treatment but an (approximate) control group exists.
- Ideas : Parallel Growth (Mora and Reggio, 2019) , Different-in-Differences with multiple time points, Synthetic Difference-in-Differences