Causal Inference

Mixtape Session

MIXTAPE SESSIONS

Roadmap

- Comparative case studies
    - Synthetic control
    - Replication exercise

Susan Athey
@Susan_Athey

Replying to @causalinf

I used it at MSFT in late 2000's and they thought it was brilliant when I explained. (I had to confess I didn't come up with the idea.) You often understand the assignment process in tech so assns are justified, just dealing with low power when units are markets, cities, etc.

2:13 PM · Feb 27, 2022 · Twitter Web App

Figure: Susan Athey, former chief economist at Microsoft and professor of economics at Stanford, on synthetic control.

What is synthetic control

- Synthetic control has been called the most important innovation in causal inference of the last 15 years (Athey and Imbens 2017)
- It's extremely useful for case studies, which is nice because that's often all we have
- Continues to also be methodologically a frontier for applied econometrics

What is a comparative case study

- Single treated unit – country, state, whatever
- Social scientists tackle such situations in two ways: qualitatively and quantitatively
- In political science, probably others, you see as a stark dividing line between camps

Qualitative comparative case studies

- In qualitative comparative case studies, the goal is to reason *inductively* the causal effects of events or characteristics of a single unit on some outcome, oftentimes through logic and historical analysis.
  - → May not answer the causal questions at all because there is rarely a counterfactual, or if so, it's ad hoc.
  - → Classic example of comparative case study approach is Alexis de Toqueville's <u>Democracy in America</u> (but he is regularly comparing the US to France)

Traditional quantitative comparative case studies

- Quantitative comparative case studies are often explicitly causal designs.
- Usually a natural experiment applied to a single aggregate unit (e.g., city, school, firm, state, country)
- Method compares the evolution of an aggregate outcome for the unit affected by the intervention to the evolution of the same *ad hoc* aggregate control group (Card 1990; Card and Krueger 1994)

Pros and cons of traditional case study approaches

- Pros:
  - → Policy interventions often take place at an aggregate level
  - → Aggregate/macro data are often available
- Cons:
  - → Selection of control group is *ad hoc*
  - → Standard errors do not reflect uncertainty about the ability of the control group to reproduce the counterfactual of interest


ELDORADO


ELDORADO


SILVER SEA
57

Description of the Mariel Boatlift

- How do inflows of immigrants affect the wages and employment of natives in local labor markets?
- Card (1990) uses the Mariel Boatlift of 1980 as a natural experiment to measure the effect of a sudden influx of immigrants on unemployment among less-skilled natives
- The Mariel Boatlift increased the Miami labor force by 7%
- Individual-level data on unemployment from the Current Population Survey (CPS) for Miami and four comparison cities (Atlanta, Los Angeles, Houston, Tampa-St. Petersburg)

Why these four?

Tables 3 and 4 present simple averages of wage rates and unemployment rates for whites, blacks, Cubans, and other Hispanics in the Miami labor market between 1979 and 1985. For comparative purposes, I have assembled similar data for whites, blacks, and Hispanics in four other cities: Atlanta, Los Angeles, Houston, and Tampa–St. Petersburg. These four cities were selected both because they had relatively large populations of blacks and Hispanics and because they exhibited a pattern of economic growth similar to that in Miami over the late 1970s and early 1980s. A comparison of employment growth rates (based on establishment-level data) suggests that economic conditions were very similar in Miami and the average of the four comparison cities between 1976 and 1984.

Card's main results

Differences-in-differences estimates of the effect of immigration on unemployment[a]

	Group	Year		
		1979 (1)	1981 (2)	1981–1979 (3)
	Whites			
(1)	Miami	5.1 (1.1)	3.9 (0.9)	− 1.2 (1.4)
(2)	Comparison cities	4.4 (0.3)	4.3 (0.3)	− 0.1 (0.4)
(3)	Difference Miami-comparison	0.7 (1.1)	− 0.4 (0.95)	− 1.1 (1.5)
	Blacks			
(4)	Miami	8.3 (1.7)	9.6 (1.8)	1.3 (2.5)
(5)	Comparison cities	10.3 (0.8)	12.6 (0.9)	2.3 (1.2)
(6)	Difference Miami-comparison	− 2.0 (1.9)	− 3.0 (2.0)	− 1.0 (2.8)

[a] Notes: Adapted from Card (1990, Tables 3 and 6). Standard errors are shown in parentheses.

Can this ever lead to subjective biases?

- Card found that the Mariel boatlift reduced unemployment *compared to the four cities he chose*
- Is there anything principled we could do that doesn't give the researcher so much control over control group?
- Enter synthetic control (Abadie and Gardeazabal 2003; Abadie, Diamond and Hainmueller 2010)

Synthetic Control

- First appears in Abadie and Gardeazabal (2003) in a study of a terrorist attack in Spain (Basque) on GDP
- Revisited again in a 2011 JASA with Diamond and Hainmueller, two political scientists who were PhD students at Harvard (more proofs and inference)
- A combination of comparison units often does a better job reproducing the characteristics of a treated unit than single comparison unit alone

Researcher's objectives

- Our goal here is to reproduce the counterfactual of a treated unit by finding the combination of untreated units that best resembles the treated unit *before* the intervention in terms of the values of k relevant covariates (predictors of the outcome of interest)
- Method selects *weighted average of all potential comparison units* that best resembles the characteristics of the treated unit(s) - called the "synthetic control"

Synthetic control method: advantages

- Precludes extrapolation (unlike regression) because counterfactual forms a convex hull
- Does not require access to post-treatment outcomes in the “design” phase of the study - no peeking
- Makes explicit the contribution of each comparison unit to the counterfactual
- Formalizing the way comparison units are chosen has direct implications for inference

Synthetic control method: disadvantages

1. Subjective researcher bias kicked down to the model selection stage
2. Significant diversity at the moment as to how to principally select models - from machine learning to modifications - as well as estimation and software

Furman and Pinto (2018) recommend showing a few different results in their "cherry picking" JPAM

Synthetic control method: estimation

Suppose that we observe $J+1$ units in periods $1, 2, \ldots, T$

- Unit “one” is exposed to the intervention of interest (that is, “treated”) during periods $T_0 + 1, \ldots, T$
- The remaining J are an untreated reservoir of potential controls (a “donor pool”)

Potential outcomes notation

- Let Y_{it}^0 be the outcome that would be observed for unit i at time t in the absence of the intervention
- Let Y_{it}^1 be the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in periods $T_0 + 1$ to T.

Dynamic ATT

Treatment effect parameter is defined as group time ATT (but with only one group) where

$$\begin{aligned}
\delta_{1t} &= Y_{1t}^1 - Y_{1t}^0 \\
&= Y_{1t} - Y_{1t}^0
\end{aligned}$$

for each post-treatment period, $t > T_0$ and Y_{1t} is the outcome for unit one at time t. We will estimate Y_{1t}^0 using the J units in the donor pool

Estimating optimal weights

- Let $W = (w_2, \ldots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{j+1} = 1$. Each value of W represents a potential synthetic control
- Let X_1 be a $(k \times 1)$ vector of pre-intervention characteristics for the treated unit. Similarly, let X_0 be a $(k \times J)$ matrix which contains the same variables for the unaffected units.
- The vector $W^* = (w_2^*, \ldots, w_{J+1}^*)'$ is chosen to minimize $||X_1 - X_0 W||$, subject to our weight constraints

Optimal weights differ by another weighting matrix

Abadie, et al. consider

$$||X_1 - X_0W|| = \sqrt{(X_1 - X_0W)'V(X_1 - X_0W)}$$

where X_{jm} is the value of the m-th covariates for unit j and V is some $(k \times k)$ symmetric and positive semidefinite matrix

More on the V matrix

Typically, V is diagonal with main diagonal $v_1, \ldots, v_k$. Then, the synthetic control weights $w_2^*, \ldots, w_{J+1}^*$ minimize:

$$\sum_{m=1}^{k} v_m \left(X_{1m} - \sum_{j=2}^{J+1} w_j X_{jm} \right)^2$$

where v_m is a weight that reflects the relative importance that we assign to the m-th variable when we measure the discrepancy between the treated unit and the synthetic controls

Choice of V is critical

- The synthetic control $W^*(V^*)$ is meant to reproduce the behavior of the outcome variable for the treated unit in the absence of the treatment
- Therefore, the V^* weights directly shape W^*

Estimating the V matrix

Choice of $v_1, \ldots, v_k$ can be based on

- Assess the predictive power of the covariates using regression
- Subjectively assess the predictive power of each of the covariates, or calibration inspecting how different values for $v_1, \ldots, v_k$ affect the discrepancies between the treated unit and the synthetic control
- Minimize mean square prediction error (MSPE) for the pre-treatment period (default):

$$\sum_{t=1}^{T_0} \left(Y_{1t} - \sum_{j=2}^{J} w_j^*(V^*) Y_{jt} \right)^2$$

Suppose Y^0 is given by a factor model

What about unmeasured factors affecting the outcome variables as well as heterogeneity in the effect of observed and unobserved factors?

$$Y_{it}^0 = \alpha_t + \theta_t Z_i + \lambda_t u_i + \varepsilon_{it}$$

where α_t is an unknown common factor with constant factor loadings across units, and λ_t is a vector of unobserved common factors.

This is the identifying assumption and like DiD it is a restriction on the counterfactual Y^0

With some manipulation

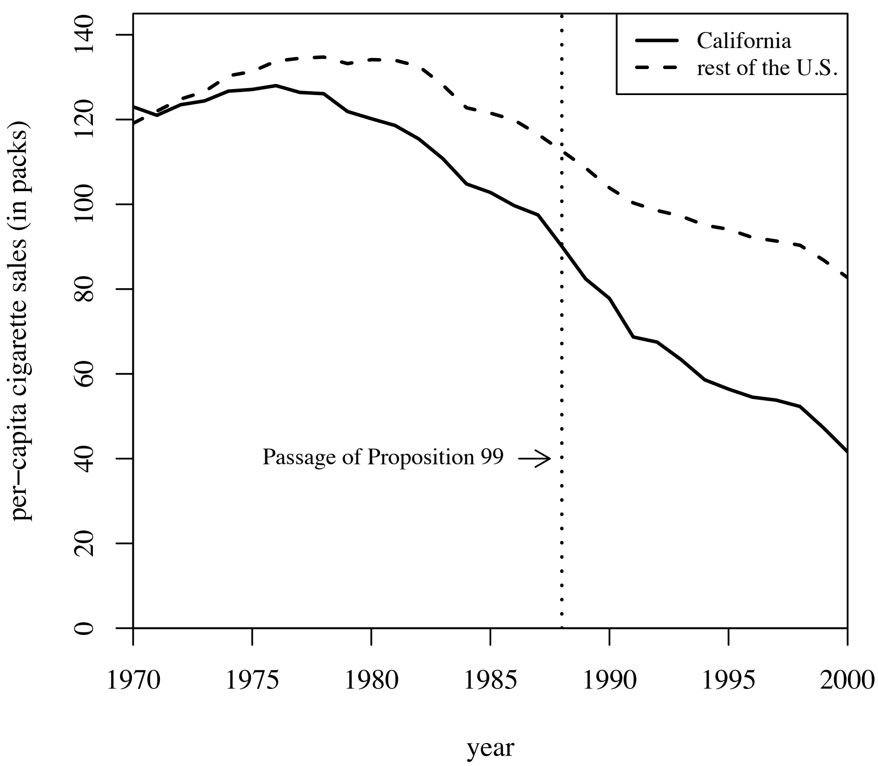
$$
\begin{aligned}
Y^0_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \quad = \quad & \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \lambda_t \left(\sum_{n=1}^{T_0} \lambda_n' \lambda_n \right)^{-1} \lambda_s' (\varepsilon_{js} - \varepsilon_{1s}) \\
& - \sum_{j=2}^{J+1} w_j^* (\varepsilon_{jt} - \varepsilon_{1t})
\end{aligned}
$$

- If $\sum_{t=1}^{T_0} \lambda_t' \lambda_t$ is nonsingular, then RHS will be close to zero if number of preintervention periods is “large” relative to size of transitory shocks
- Only units that are alike in observables and unobservables should produce similar trajectories of the outcome variable over extended periods of time
- Proof in Appendix B of ADH (2011)

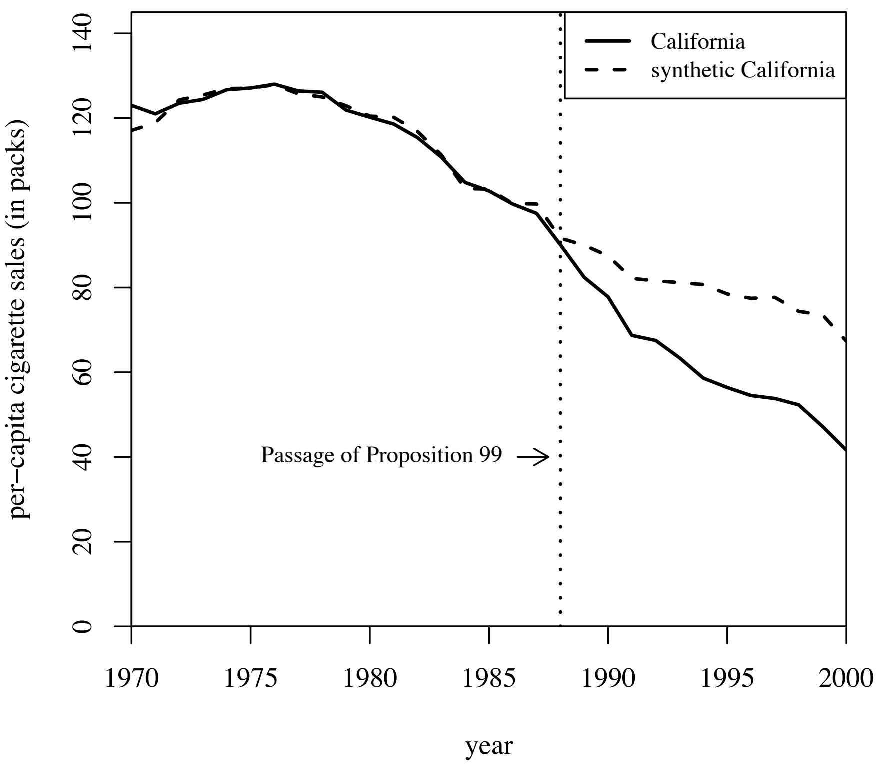
Example: California's Proposition 99

- In 1988, California first passed comprehensive tobacco control legislation:
  - → increased cigarette tax by 25 cents/pack
  - → earmarked tax revenues to health and anti-smoking budgets
  - → funded anti-smoking media campaigns
  - → spurred clean-air ordinances throughout the state
  - → produced more than $100 million per year in anti-tobacco projects
- Other states that subsequently passed control programs are excluded from donor pool of controls (AK, AZ, FL, HI, MA, MD, MI, NJ, OR, WA, DC)

Cigarette Consumption: CA and the Rest of the US



Cigarette Consumption: CA and synthetic CA

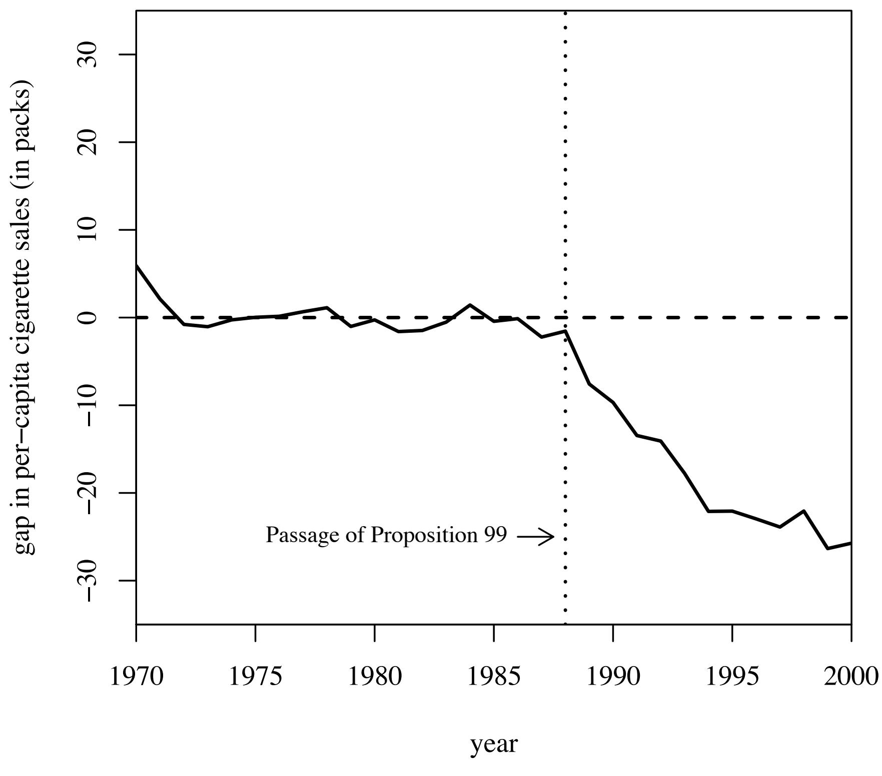


Predictor Means: Actual vs. Synthetic California

	California		Average of
Variables	Real	Synthetic	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).

Smoking Gap between CA and synthetic CA



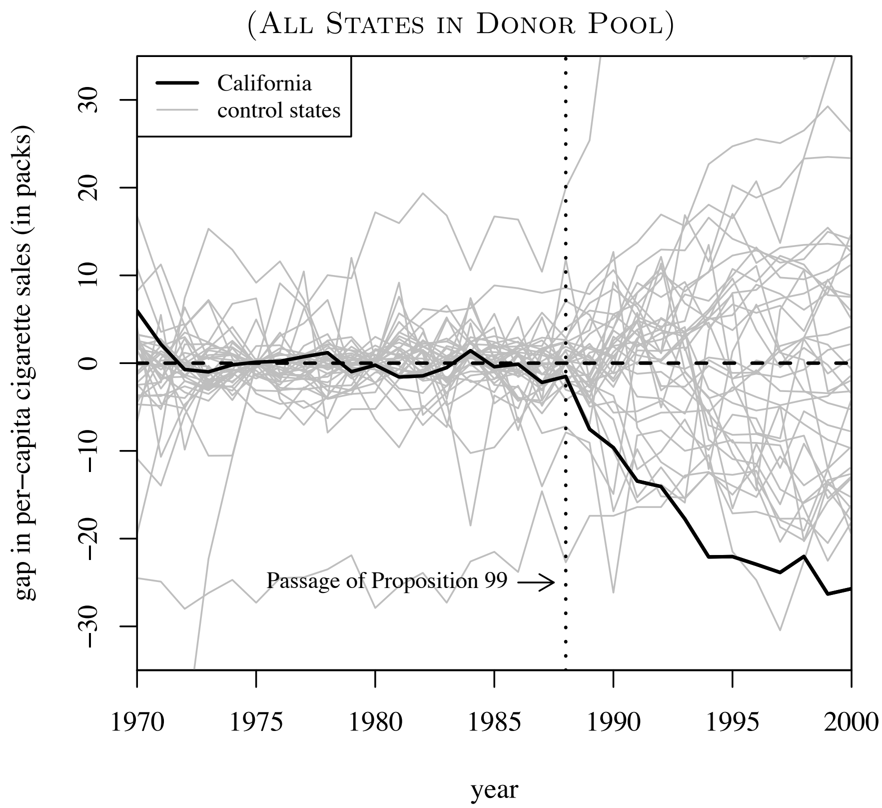
Inference

- To assess significance, we calculate exact p-values under Fisher's sharp null using a test statistic equal to after to before ratio of RMSPE
- Exact p-value method
  - → Iteratively apply the synthetic method to each country/state in the donor pool and obtain a distribution of placebo effects
  - → Compare the gap (RMSPE) for California to the distribution of the placebo gaps. For example the post-Prop. 99 RMSPE is:

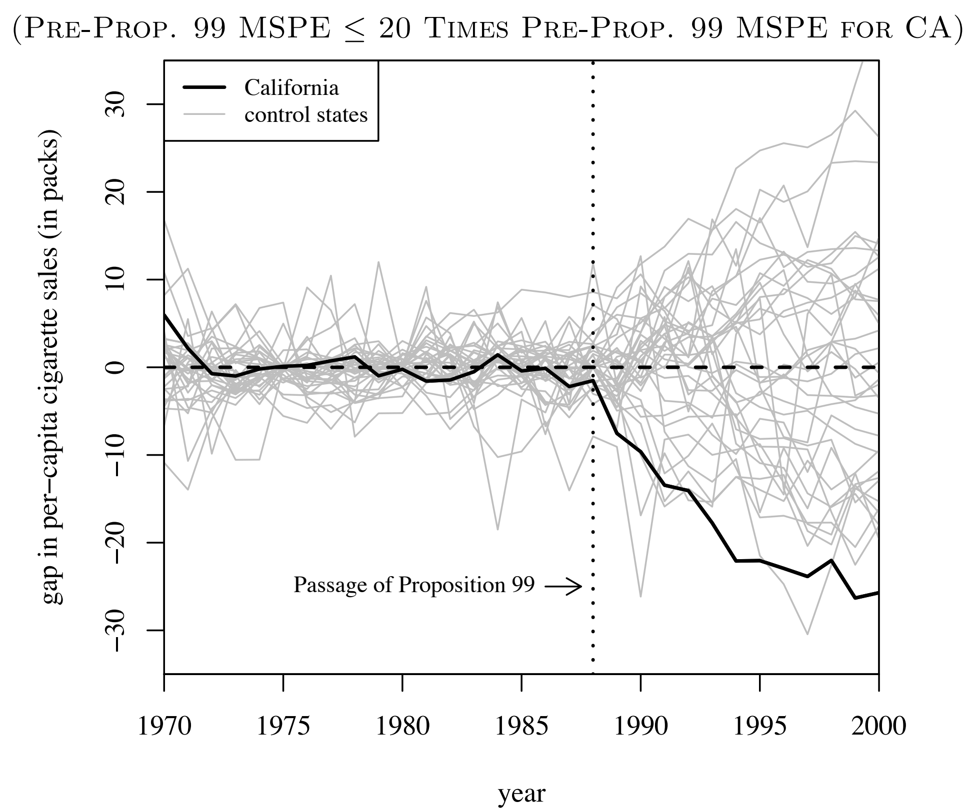
$$RMSPE = \left(\frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \right)^2 \right)^{\frac{1}{2}}$$

and the exact p-value is the treatment unit rank divided by J

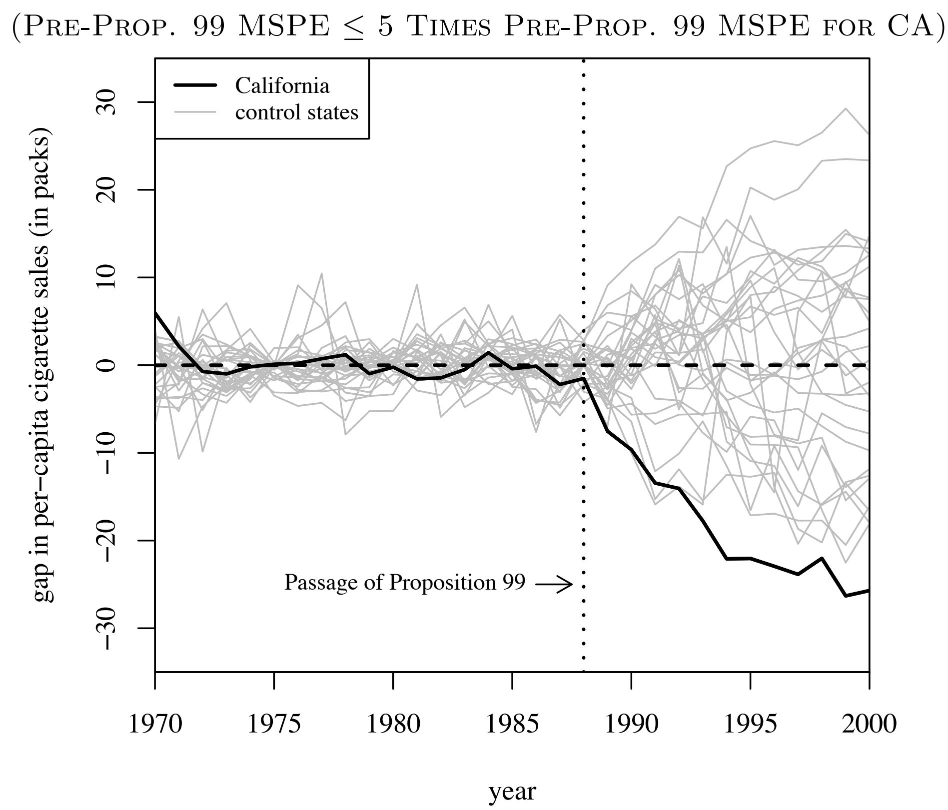
Smoking Gap for CA and 38 control states



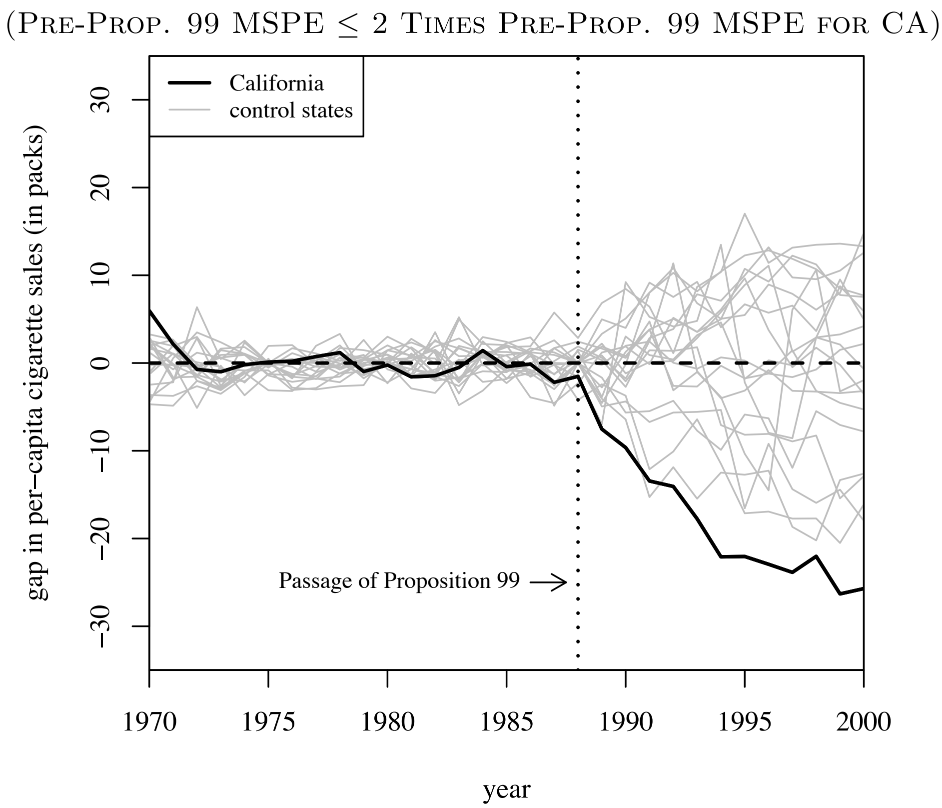
Smoking Gap for CA and 34 control states



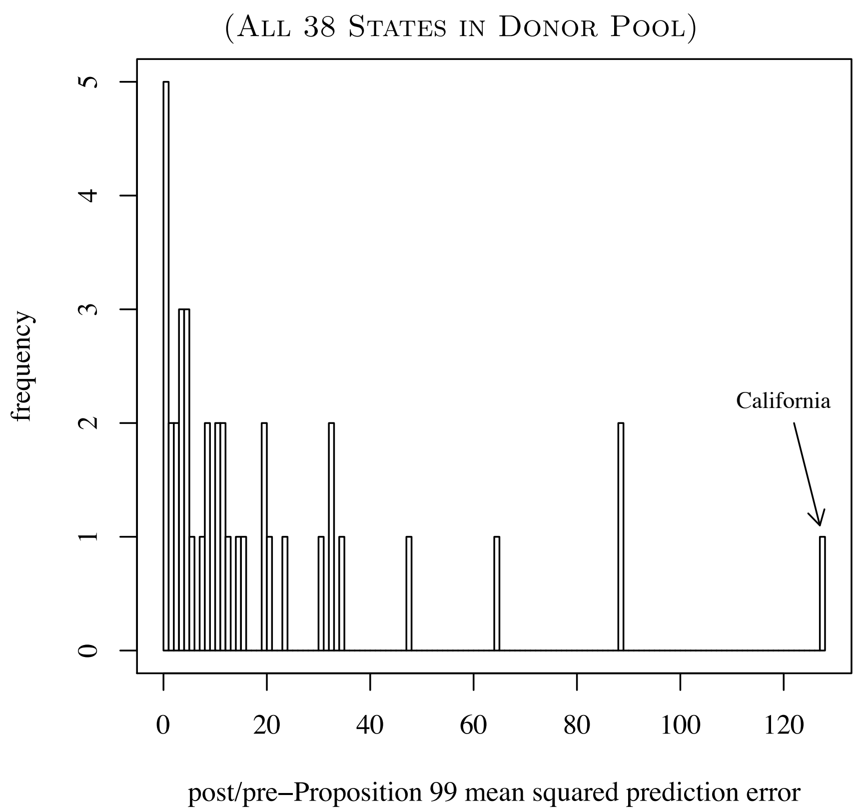
Smoking Gap for CA and 29 control states



Smoking Gap for CA and 19 control states



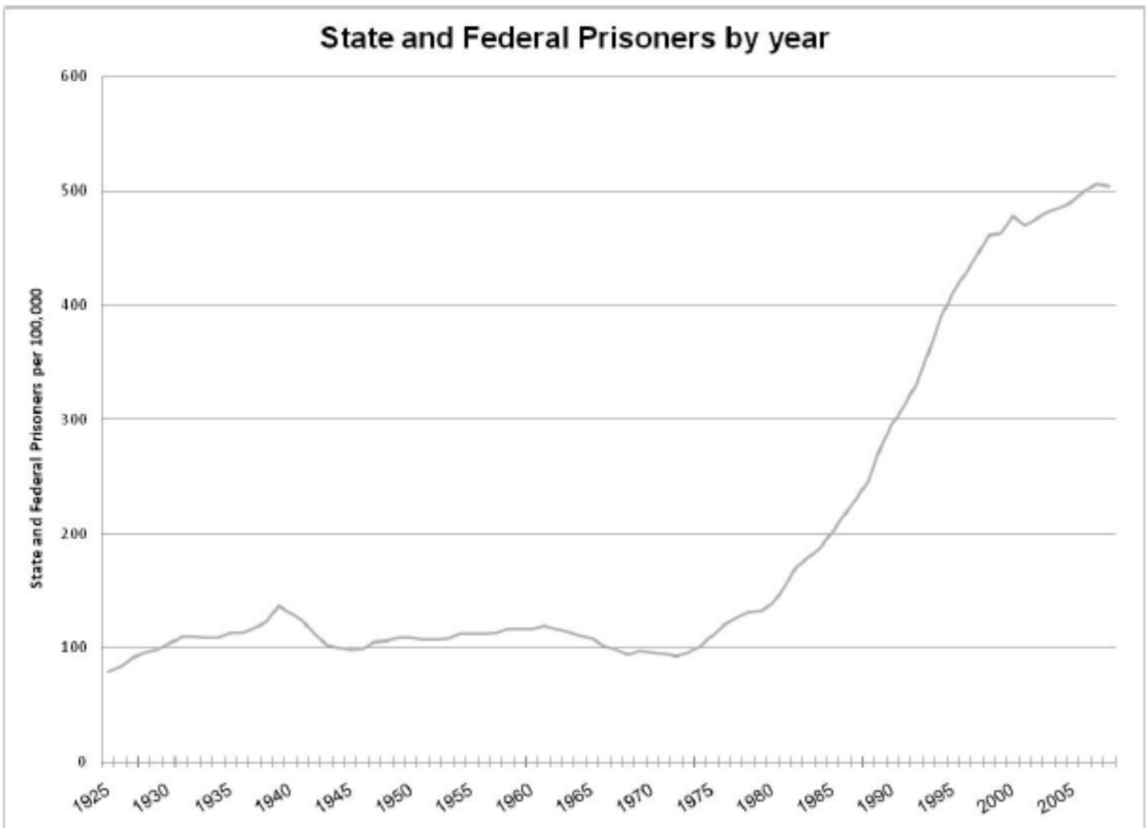
Ratio Post-Prop. 99 RMSPE to Pre-Prop. 99 RMSPE



Facts

- The US has the highest prison population of any OECD country in the world
- 2.3 million are currently incarcerated in US federal and state prisons and county jails
- Another 4.75 million are on parole
- From the early 1970s to the present, incarceration and prison admission rates quintupled in size

Figure 1
History of the imprisonment rate, 1925 - 2008



Source: www.albany.edu/sourcebook/tost_6/html

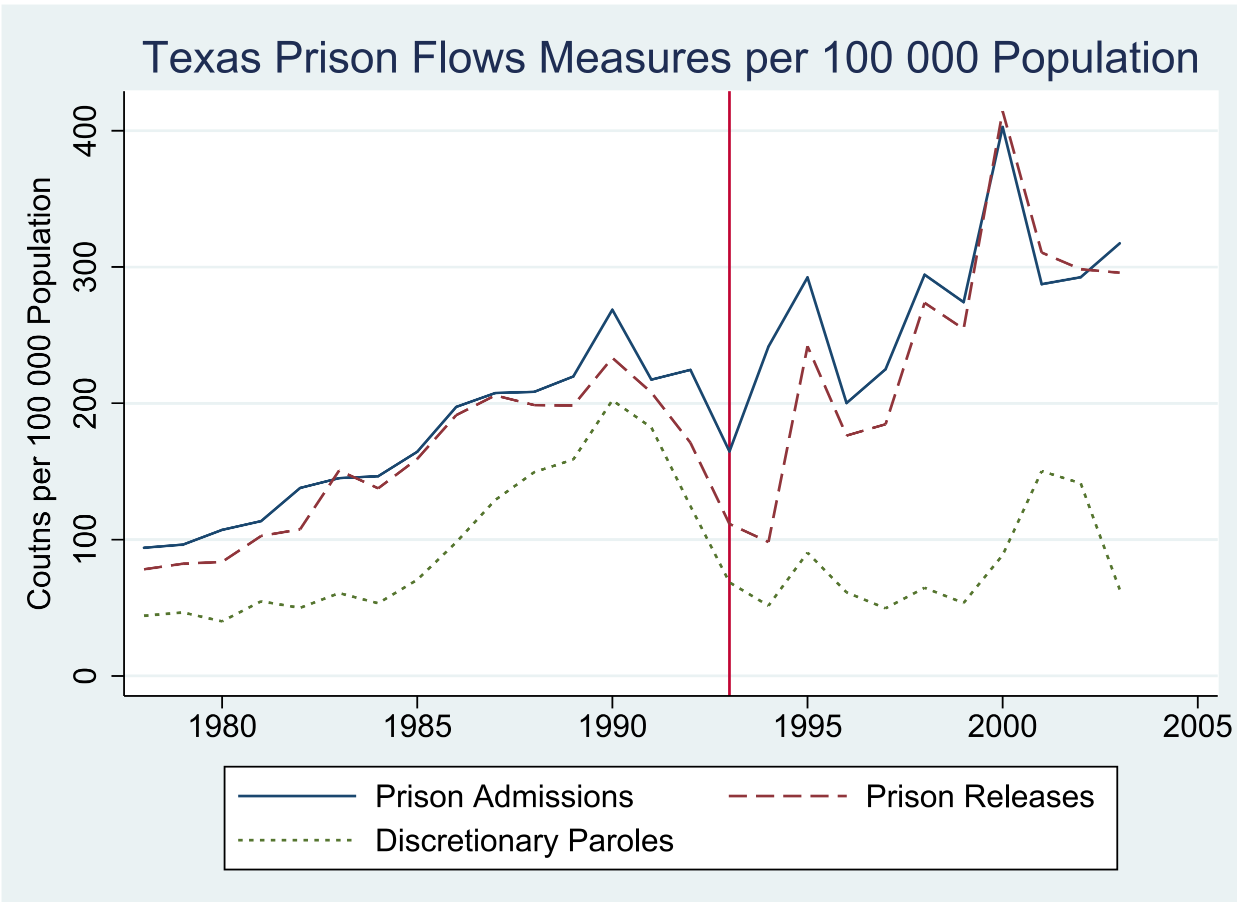
TDC lawsuit

Ruiz v. Estelle 1980

- Class action lawsuit against TX Dept of Corrections (Estelle, warden).
- TDC lost. Lengthy period of appeals and legal decrees.
- Lengthy period of time relying on paroles to manage flows

Prison constraints

- Prisons are and have been at capacity for a long time.
- Requires managing flows through
  - → Prison construction
  - → Overcrowding
  - → Paroles

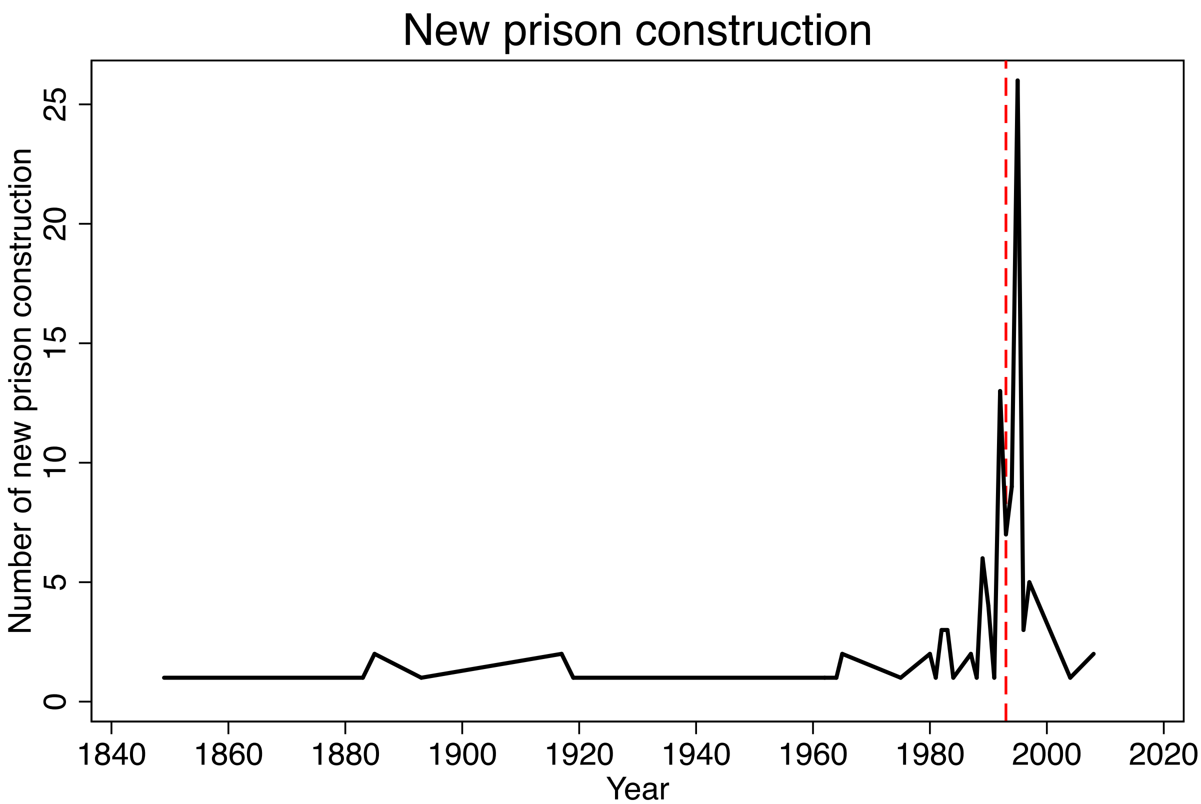


Texas Prison Flows Measures per 100 000 Population
Coutns per 100 000 Population
0
100
200
300
400
1980
1985
1990
1995
2000
2005
Prison Admissions
Prison Releases
Discretionary Paroles

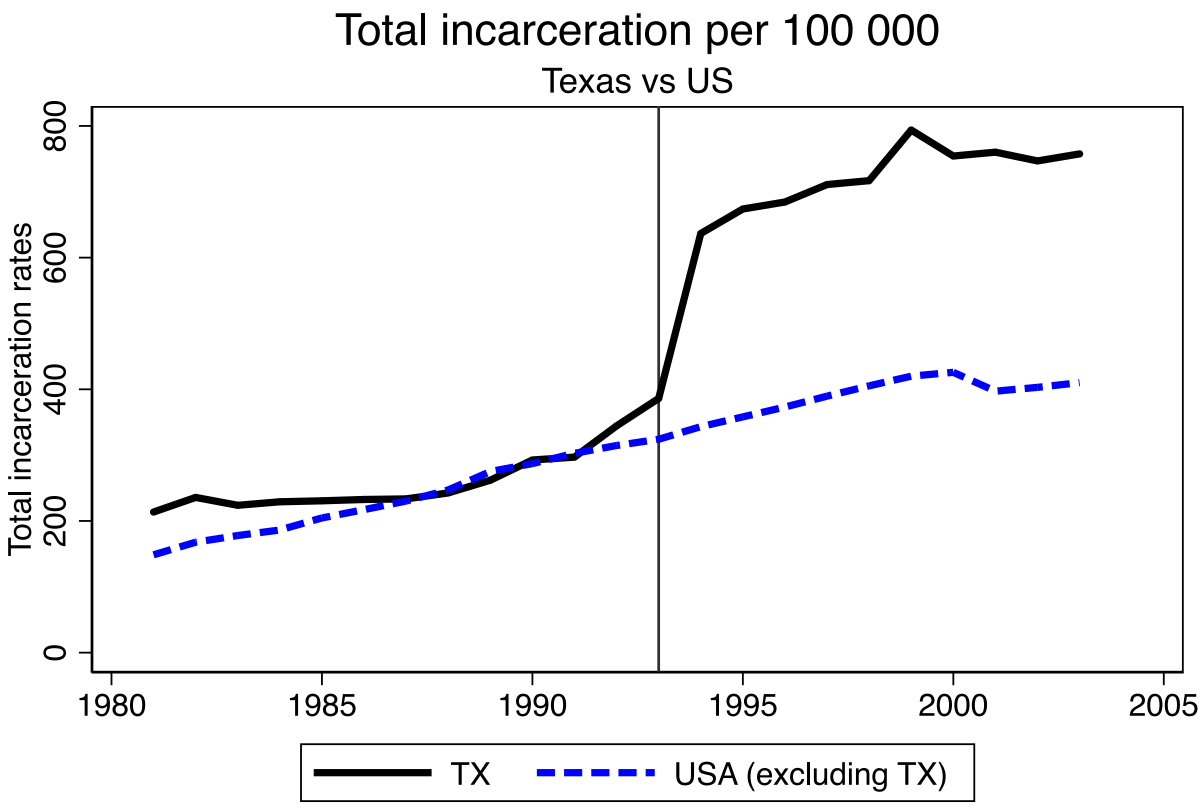
Prison construction

Governor Ann Richards (D) 1991-1995 raises a billion dollars to expand cells throughout the state

- Operation prison capacity increased 30-35% in 1993, 1994 and 1995.
- Prison capacity increased from 55,000 in 1992 to 130,000 in 1995.
- Building of new prisons (private and public)



Red dashed line is 1993



1993 starts the prison expansion