Synthetic Control: Part II

INFO/STSCI/ILRST 3900: Causal Inference

13 Nov 2025

Learning goals for today

At the end of class, you will be able to:

1. Understand example of synthetic control applied to anti-smoking legislation
2. Understand how to conduct a hypothesis test for estimates from a synthetic control analysis
3. Reason about how interference will affect estimates from a synthetic control analysis

Logistics

▶ This week, read Ch 10 of The Causal Inference Mixtape
▶ Problem Set 5 peer reviews due Nov 18
▶ Quiz 5 Nov 18
▶ Discussion sessions next week will be project help sessions
▶ Project Check-ins due Nov 25
▶ Lectures on interference next week

Example: Smoking in California

▶ *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program* by Abadie, Diamond, and Hainmueller (2012)

Example: Smoking in California

- *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program* by Abadie, Diamond, and Hainmueller (2012)
- In the 1950 and 1960s, the public began realizing that smoking posed health risks
- In 1988, California passed prop 99 which increased taxes on cigarettes by 25 cents
- Earmarked funds for anti-smoking campaigns and research

Example: Smoking in California

- ▶ *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program* by Abadie, Diamond, and Hainmueller (2012)
- ▶ In the 1950 and 1960s, the public began realizing that smoking posed health risks
- ▶ In 1988, California passed prop 99 which increased taxes on cigarettes by 25 cents
- ▶ Earmarked funds for anti-smoking campaigns and research
- ▶ Some measures were rolled back in early 90s, but other measures ensured smoke free restaurants and workplaces
- ▶ Similar measures enacted in Massachusetts, Oregon, Arizona...

Example: Smoking in California

- ▶ *Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program* by Abadie, Diamond, and Hainmueller (2012)
- ▶ In the 1950 and 1960s, the public began realizing that smoking posed health risks
- ▶ In 1988, California passed prop 99 which increased taxes on cigarettes by 25 cents
- ▶ Earmarked funds for anti-smoking campaigns and research
- ▶ Some measures were rolled back in early 90s, but other measures ensured smoke free restaurants and workplaces
- ▶ Similar measures enacted in Massachusetts, Oregon, Arizona...

What was the effect of Proposition 99 (and the related measures) on smoking in California?

Synthetic Control: big idea

- ▶ Data have many pre- and post-treatment periods
- ▶ Treated unit is "unique"
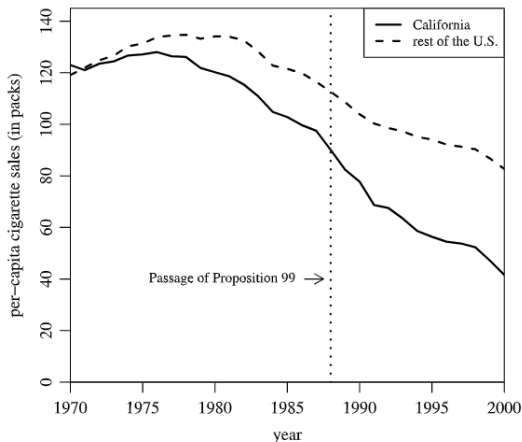- ▶ Not so many units in control group

Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Synthetic Control: big idea

- ▶ Data have many pre- and post-treatment periods
- ▶ Treated unit is "unique"
- ▶ Not so many units in control group
- ▶ Construct synthetic unit to approximate untreated version of treated unit using weighted average of untreated units
- ▶ Pick weights to match pre-treatment characteristics (either covariates or observations)

$$\underbrace{\sum_{t < T_0}}_{pre-treatment} \left(Y_{t,CA} - \underbrace{\sum_j w_j Y_{t,j}}_{Synthetic\ CA} \right)^2$$

- ▶ The weights are non-negative and sum to 1 so the Synthetic CA is interpretable

Example: Smoking in California

▶ Outcome: cigarette sales per capita
▶ Treatment: Prop 99, increase in cigarette tax and anti-smoking campaign
▶ Donor pool consists of 38 states which did not enact increase in cigarette taxes
  ▶ If we were to use "regular" matching, would we be able to find a good match for California?
  ▶ If we were to use diff-in-diff, should we use average of all 38 other states?

Example: Smoking in California

Table 2. State weights in the synthetic California

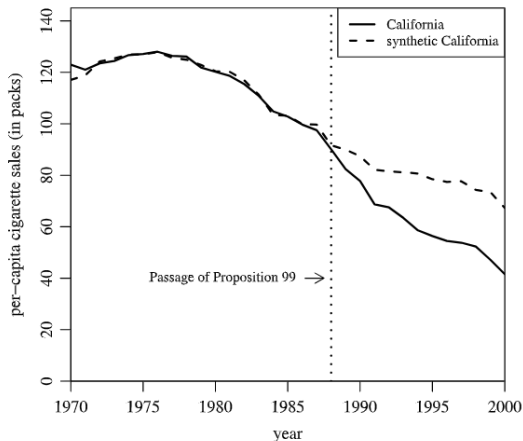| State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	–	Nebraska	0
Arizona	–	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	–
Connecticut	0.069	New Mexico	0
Delaware	0	New York	–
District of Columbia	–	North Carolina	0
Florida	–	North Dakota	0
Georgia	0	Ohio	0
Hawaii	–	Oklahoma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

Example: Smoking in California

Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Example: Smoking in California

Table 1. Cigarette sales predictor means

Variables	California		Average of 38 control states
	Real	Synthetic	
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.

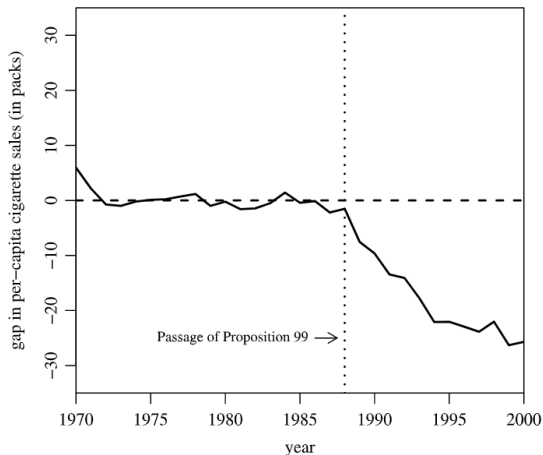# Example: Smoking in California

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

Hypothesis testing

Hypothesis testing

- ▶ In general, we expect a discrepancy between the synthetic unit and the real observations, even if the treatment had no effect
- ▶ How might we test whether the estimated effect is statistically significant?

Hypothesis testing

- In general, we expect a discrepancy between the synthetic unit and the real observations, even if the treatment had no effect
- How might we test whether the estimated effect is statistically significant?
- Placebo test:
    - Run synthetic control on other states which didn't get treatment
    - Compare difference between observed and synthetic California to the difference between observed and synthetic for untreated states
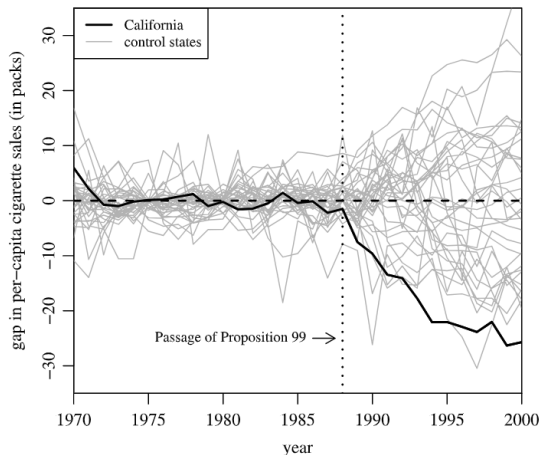
Example: Smoking in California

Figure 5. Per-capita cigarette sales gaps in California and placebo gaps in 34 control states (discards states with pre-Proposition 99 MSPE twenty times higher than California's).

Hypothesis testing

- Discrepancy for California:

$$\text{Ratio of Mean Squared Error} = \frac{\sum_{t \geq T_0}(Y_{t,CA} - Y_{t,synthCA})^2}{\sum_{t < T_0}(Y_{t,CA} - Y_{t,synthCA})^2}$$

Hypothesis testing

- Discrepancy for California:

$$\text{Ratio of Mean Squared Error} = \frac{\sum_{t \geq T_0}(Y_{t,CA} - Y_{t,synthCA})^2}{\sum_{t < T_0}(Y_{t,CA} - Y_{t,synthCA})^2}$$

- Compare to RMSE for untreated states
- If California is large relative to other states, indicates difference is unlikely to occur simply by chance

Hypothesis testing

- Discrepancy for California:

  $$\text{Ratio of Mean Squared Error} = \frac{\sum_{t \geq T_0}(Y_{t,CA} - Y_{t,synthCA})^2}{\sum_{t < T_0}(Y_{t,CA} - Y_{t,synthCA})^2}$$

- Compare to RMSE for untreated states
- If California is large relative to other states, indicates difference is unlikely to occur simply by chance
- P-value: is proportion of states larger than California
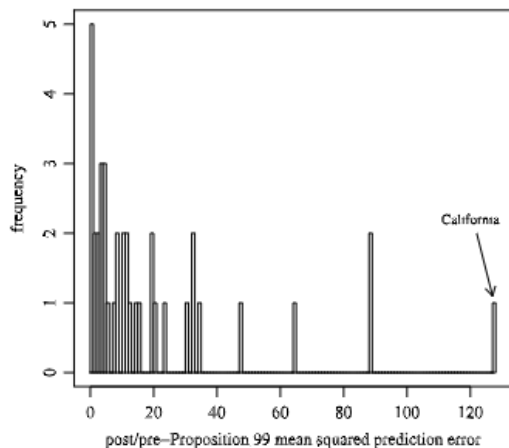
Exampl: Smoking in California

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

P-value = .026

Interference

Did treatment in California also affect other states?

- ▶ Perhaps Prop 99 in California increase anti-smoking sentiment in other states?
- ▶ If so, how would this affect our estimated effect size?

Interference

Did treatment in California also affect other states?

- $Y_{t,j}^0$ is the cigarettes purchased per capita in state j in a world where no anti-smoking legislation passed (Prop 99 or otherwise)

- Suppose $Y_{t,CA}^0 = \sum_j w_j Y_{t,j}^0$

Interference

Did treatment in California also affect other states?

- $Y_{t,j}^0$ is the cigarettes purchased per capita in state j in a world where no anti-smoking legislation passed (Prop 99 or otherwise)

- Suppose $Y_{t,CA}^0 = \sum_j w_j Y_{t,j}^0$

- If after T_0 we have $Y_{t,j} = Y_{t,j}^0$, then

$$Y_{t,Synth} = \sum_j \hat{w}_j Y_{t,j} = \sum_j \hat{w}_j Y_{t,j}^0 \approx \sum_j w_j Y_{t,j}^0 = Y_{t,CA}^0, \quad and$$

$$Y_{t,CA} - Y_{t,Synth} = Y_{t,CA} - \sum_j \hat{w}_j Y_{t,j} \approx Y_{t,CA}^1 - Y_{t,CA}^0$$

Interference

Did treatment in California also affect other states?

- $Y_{t,j}^0$ is the cigarettes purchased per capita in state j in a world where no anti-smoking legislation passed (Prop 99 or otherwise)

- Suppose $Y_{t,CA}^0 = \sum_j w_j Y_{t,j}^0$

- If after T_0 we have $Y_{t,j} = Y_{t,j}^0$, then

$$Y_{t,Synth} = \sum_j \hat{w}_j Y_{t,j} = \sum_j \hat{w}_j Y_{t,j}^0 \approx \sum_j w_j Y_{t,j}^0 = Y_{t,CA}^0, \quad and$$

$$Y_{t,CA} - Y_{t,Synth} = Y_{t,CA} - \sum_j \hat{w}_j Y_{t,j} \approx Y_{t,CA}^1 - Y_{t,CA}^0$$

- If after T_0 we have $Y_{t,j} = Y_{t,j}^0 - 5$, then

$$Y_{t,Synth} = \sum_j \hat{w}_j Y_{t,j} = \sum_j \hat{w}_j (Y_{t,j}^0 - 5) \approx \sum_j w_j (Y_{t,j}^0 - 5) = Y_{t,CA}^0 - 5, \quad and$$

$$Y_{t,CA} - Y_{t,Synth} = Y_{t,CA} - \sum_j \hat{w}_j Y_{t,j} \approx Y_{t,CA}^1 - (Y_{t,CA}^0 - 5) = Y_{t,CA}^1 - Y_{t,CA}^0 + 5$$

Interference

Did treatment in California also affect other states?

Interference

Did treatment in California also affect other states?

- ▶ Did increased taxes on cigarettes cause Californians to buy cigarettes in other states?
- ▶ If so, how would this affect our estimated effect size?
- ▶ Did Californians anticipate higher tax on cigarettes in 1988 and stock up in 1987?
- ▶ If so, how would this affect our estimated effect size?

Learning goals for today

At the end of class, you will be able to:

1. Understand example of synthetic control applied to anti-smoking legislation

2. Understand how to conduct a hypothesis test for estimates from a synthetic control analysis

3. Reason about how interference will affect estimates from a synthetic control analysis