Synthetic Control Method

Prof. Tzu-Ting Yang
楊子霆

Institute of Economics, Academia Sinica
中央研究院經濟研究所

April 17, 2023

Synthetic Control Method: Main Idea

Main Idea of Synthetic Control Method

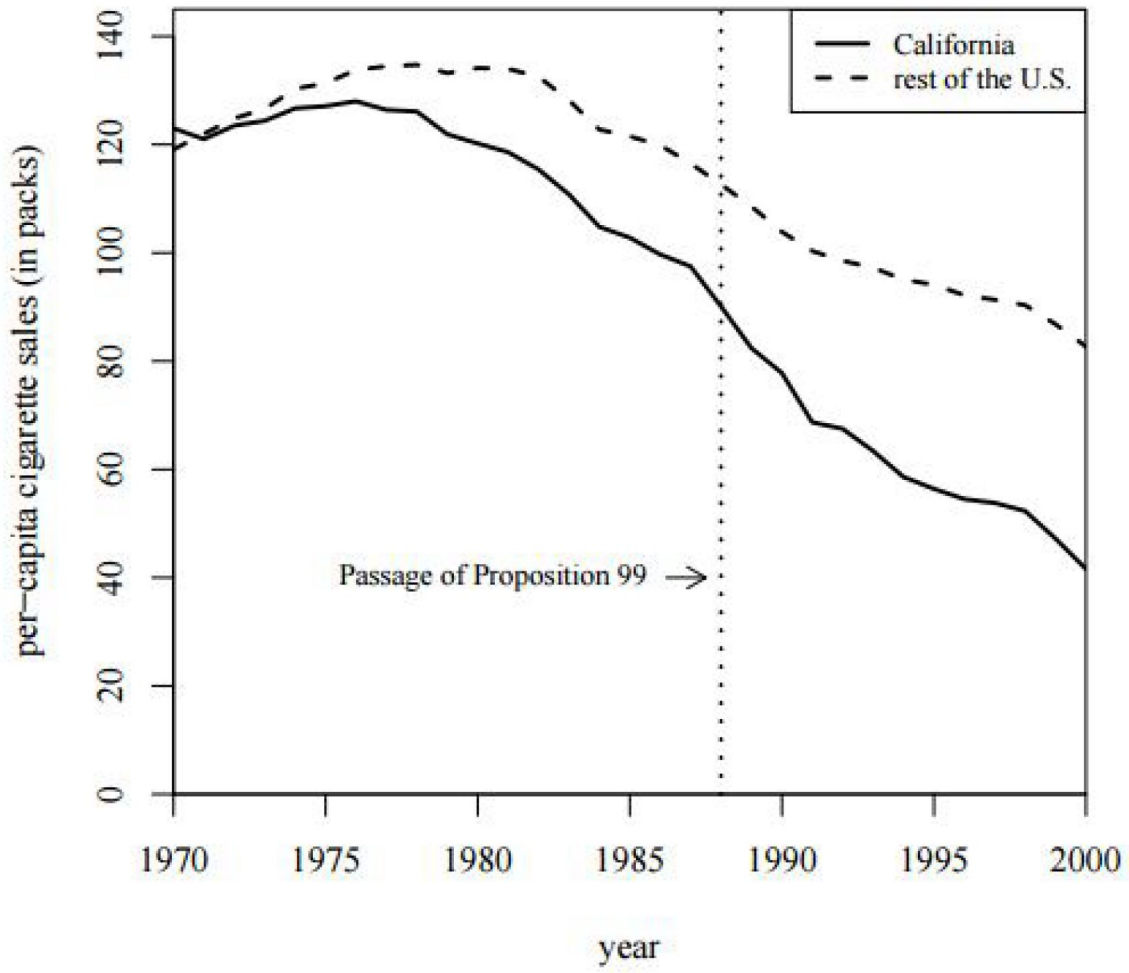
- Synthetic Control (SC) is a method to evaluate the causal effect of treatment.
- SC is quite popular in social science due to the following features:
  1. It can evaluate treatment effects on aggregate outcomes of **one (or very few) treated unit** (e.g. county-level crime rate)
  2. It is a **data-driven procedure**
  3. Use **a small number of non-treated units** to build the suitable counterfactuals

Main Idea of Synthetic Control Method

- Main idea:
  - Use (long) panel data to build the **weighted average of non-treated units**
    - The **weighted average of non-treated units** is the **synthetic unit**
    - Synthetic unit can best reproduce characteristics of the **treated unit** over time in pre-treatment period
  - Causal effect of treatment can be quantified by:
    - A simple difference in the post-treatment period: **treated unit vs synthetic unit**

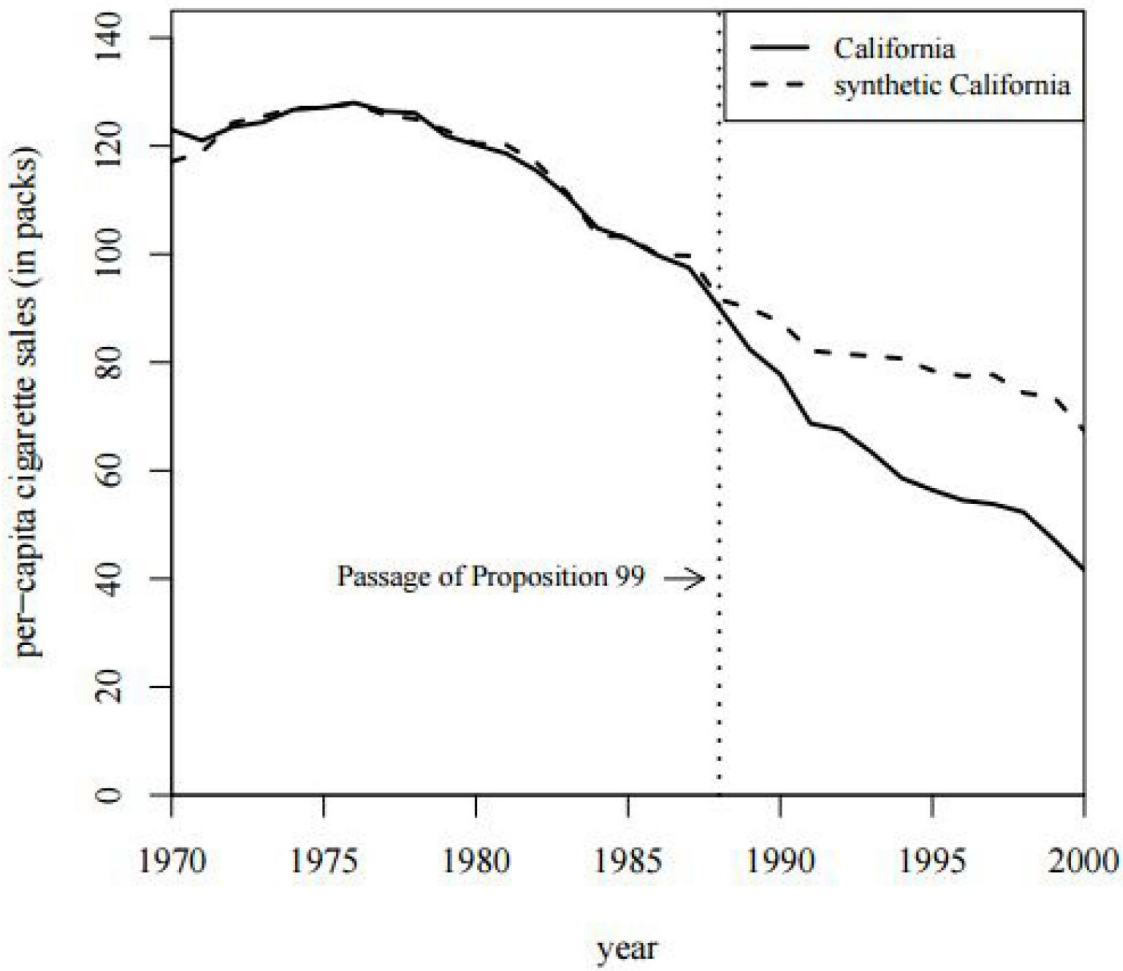
Main Idea of Synthetic Control Method

Graphical Representation



Main Idea of Synthetic Control Method

Graphical Representation



SC Method v.s. DID Method

Advantages

- SC builds upon the setting of the standard DID model, but makes the following changes:
  - SC takes a serious, data driven approach to forming counterfactuals/selecting the control group
    - Data-driven procedures reduce discretion in the choice of the control group
  - SC allows for time-varying individual-specific heterogeneity

SC Method v.s. DID Method

Disadvantages

- The benefits of SC method come with three costs:
  1. Large scale (asymptotic) inference cannot be conducted on SC estimators
     - Instead, SC uses **permutation method** to do statistical inference (compute p-value)
  2. A key identification assumption is somewhat ambiguous
  3. Computational time far exceeds that of DID estimators

Synthetic Control Method: Potential Outcomes Framework

Basic Setup: Single Treated Model

Example

Alberto Abadie, 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

- This is one of the first paper using SC method
  - Use this example to go through the key concepts of SC method
  - Apply SC method to study the effects of Proposition 99, a large-scale tobacco control program that California implemented in 1988
  - A single treated unit with multiple non-treated units

Basic Setup: Single Treated Model

Example

- In 1988, California passed comprehensive tobacco control legislation:
  - Increased cigarette taxes by \$0.25 per pack
  - Funded anti-smoking media campaigns
  - Spurred clean-air ordinances
- They want to estimate the causal effect of the policy on cigarette consumption in California

Basic Setup: Single Treated Model

- Suppose we observe $J+1$ units over $t = 1, ..., T$ periods
- A "treatment" occurs at period $T_0 + 1$
  - Unit 1 being treated
  - Units $\{2, ..., J+1\}$ being unaffected
  - Pre-treatment period: $1 ..... T_0$
  - Post-treatment period: $T_0 + 1 ..... T$
- We aim to measure the causal effect of the treatment on the treated unit 1

SC and Potential Outcomes

- Treatment
  - $D_{it} = 1$: the units that are treated from periods $T_0 + 1$ until T
  - $D_{it} = 0$: the units that are always untreated

SC and Potential Outcomes

- Potential Outcomes
  - Y_{it}^{1}: the potential outcome we *would* observe for unit i at time t if unit i receives the treatment
    - Note that the treated unit would receive treatment from periods $T_0 + 1$ until T
  - Y_{it}^{0}: the potential outcome we *would* observe for unit i at time t if unit i does not receives the treatment
- Note that **unit in synthetic control method is usually aggregate level**: country, state, county, or region

SC and Potential Outcomes

- Observed Outcomes
  - Y_{it} is the observed outcome for unit i at time t
    - Observed outcomes before period $T_0 + 1$:

$$Y_{it} = \mathrm{Y}_{it}^{0}$$

    - Observed outcomes after period $T_0 + 1$:

$$Y_{it} = \mathrm{Y}_{it}^{0}(1 - D_{it}) + \mathrm{Y}_{it}^{1} D_{it}$$

SC and Potential Outcomes

- Since only unit 1 is treated, we aim to estimate the causal effect of treatment over time $(T_{0+1}, ..., T)$ for the treated unit 1

$$\alpha_{1t} = (\alpha_{1T_0+1}, ..., \alpha_{1T})$$

where for $t > T_0$:

$$\alpha_{1t} = Y^1_{1t} - Y^0_{1t} = \underbrace{Y^1_{1t}}_{\text{observed}} - \underbrace{Y^0_{1t}}_{\text{counterfactual}}$$

- Therefore, we need to construct the **unobserved counterfactual**

Assumptions of SC

- Assume potential outcome if unit 1 would not receive treatment Y_{1t}^0 can be expressed as follows:

$$Y_{1t}^0 = \delta_t + \theta_t Z_1 + \lambda_t \mu_1 + \varepsilon_{1t}$$

  - δ_t are common time effects (e.g. year fixed effects)
  - Z_1 are the observed, pre-treatment covariates
  - μ_1 are **permanent** unobserved variables
  - ε_{1t} are unobserved transitory shocks at the unit level with zero mean

- This assumption allows time-varying responsses of multiple unobserved factors ($\lambda_t \mu_1$)
- However, it implicitly assumes **the number of unobservable factors μ_1 are fixed over the period**
  - No structural breaks

Synthetic Control Method: Estimation

SC Estimation

$$\alpha_{1t} = Y^1_{1t} - Y^0_{1t} = \underbrace{Y^1_{1t}}_{\text{observed}} - \underbrace{Y^0_{1t}}_{\text{counterfactual}}$$

- SC method suggests treatment effect can be estimated by the simple difference:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{i=2}^{J+1} w_i^* Y_{it}$$

SC Estimation

- Choose $W = (w_2^*, ..., w_{J+1}^*) \in [0,1]$ to minimize difference in pre-treatment characteristics X between treated and weighted average of non-treated units
  - Minimize $||X_1 - X_i W||$
  - X includes observed characteristics Z and pre-treatment outcomes $Y_1, ... Y_{T0}$
  - subject to $\sum_{i=2}^{J+1} w_i^* = 1$
- Thus, different weights W gives different synthetic units

SC Estimation

- Ideally, one would want to select W^* such that

$$\sum_{j=2}^{J+1} w_i^* Z_i = Z_1$$

$$\sum_{j=2}^{J+1} w_i^* \mu_i = \mu_1$$

- Thus, causal effect of treatment $\hat{\alpha}_{1t}$ is unbiased

SC Estimation

- Problem: μ_i **is unobserved**

SC Estimation

- Solution: Choose $W = (w_1^*, ..., w_J^*)$ satisfying

$$\sum_{i=2}^{J+1} w_i^* Z_i = Z_1,$$

$$\sum_{i=2}^{J+1} w_i^* Y_{i1} = Y_{11},$$

$$\sum_{i=2}^{J+1} w_i^* Y_{i2} = Y_{12}, ...,$$

$$\sum_{i=2}^{J+1} w_i^* Y_{iT_0} = Y_{1T_0}$$

SC Estimation

- **Theorem (SC Estimation)**: Suppose there exists W^* such that the SC matches the treated unit in the pre-treatment period:

$$\sum_{j=2}^{J+1} w_i^* Y_{it} = Y_{1t} \; \forall t \in \{1, ..., T_0\}$$

$$\sum_{j=2}^{J+1} w_i^* Z_j = Z_1,$$

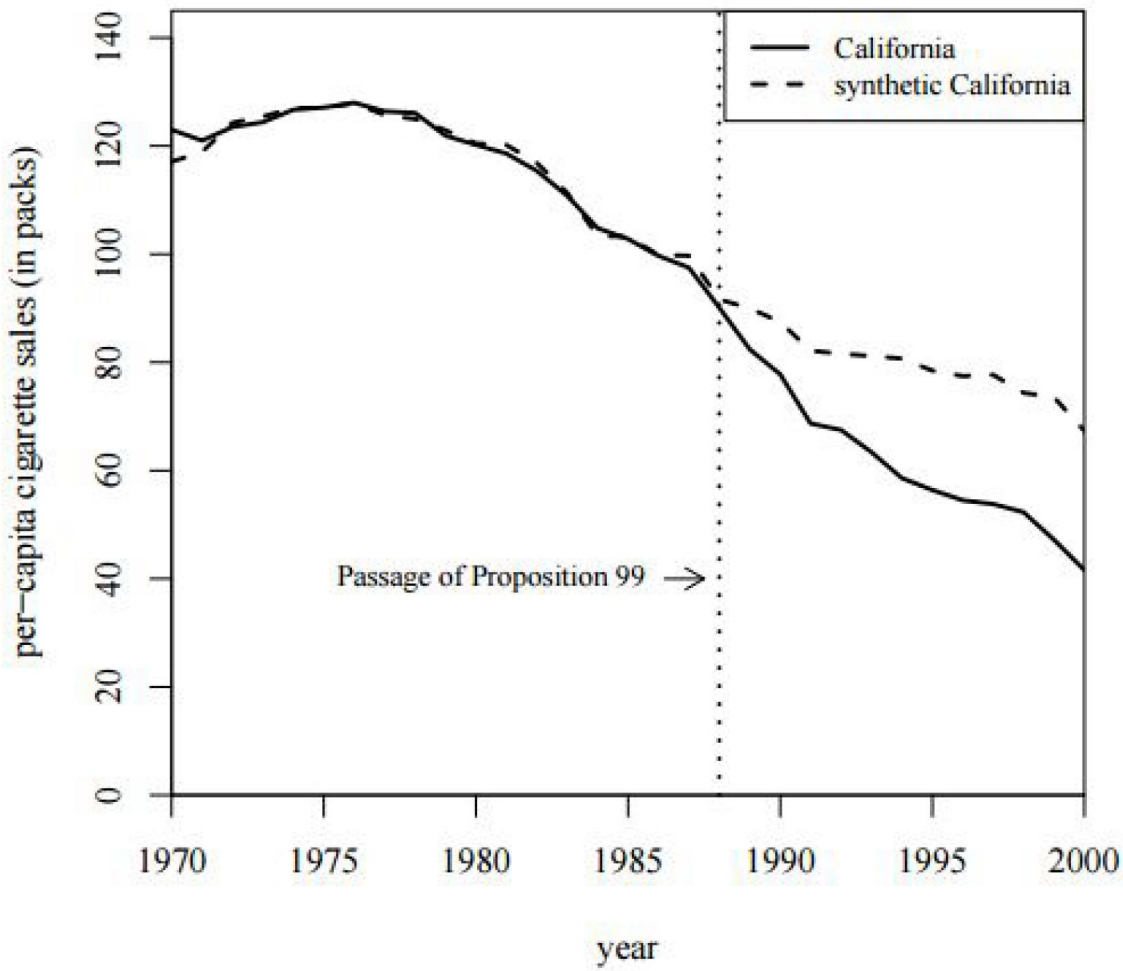
and $\sum_{t=1}^{T_0} \lambda_t' \lambda_t$ is non-singluar. Then for all $t > T_0$ we have

$$\mathbb{E}\left[Y_{1t}^0 - \sum_{i=2}^{J+1} w_i^* Y_{it} \right] \to 0$$

as $T_0 \to \infty$

Main Idea of Synthetic Control Method

Graphical Representation



SC Estimation: Intuition

- An approximately unbiased estimator of causal effect of treatment α_{1t} is then given by:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{i=2}^{J+1} w_i^* Y_{it}$$

  - The difference between the observed outcome and the outcome of synthetic unit.

- Beauty of SC is that even though the μ_i **are unobservable**, fitting $\{Y_{it}, Z_i\}$ is sufficient to match the process

- **Intuition:** a synthetic cohort can fit $(Z_1, Y_{11}, \ldots Y_{1T_0})$ for a large T_0 only if it fits (Z_1, μ_1)

Empirical Example 1: Abadie et. al (2010)

Results: Unit Weights for SC

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

Empirical Example 1: Abadie et. al (2010)

Comparison of Synthetic Fit and Simple Average

	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

Statistical Inference of SC

- Large-sample asymptotic inference is not possible with SC
- Abadie et. al (2010) suggest the use of **permutation methods** for inference
- **Main Idea:**
  - How often would we obtain results of this magnitude if we had chosen a state at random for the study instead of California ?
  - Run placebo studies by applying the SC method to states that did NOT receive treatment
    - If the placebo studies create gaps of magnitude similar to the one estimated for California
    - Then our interpretation is that our analysis does NOT provide significant evidence of causal effect

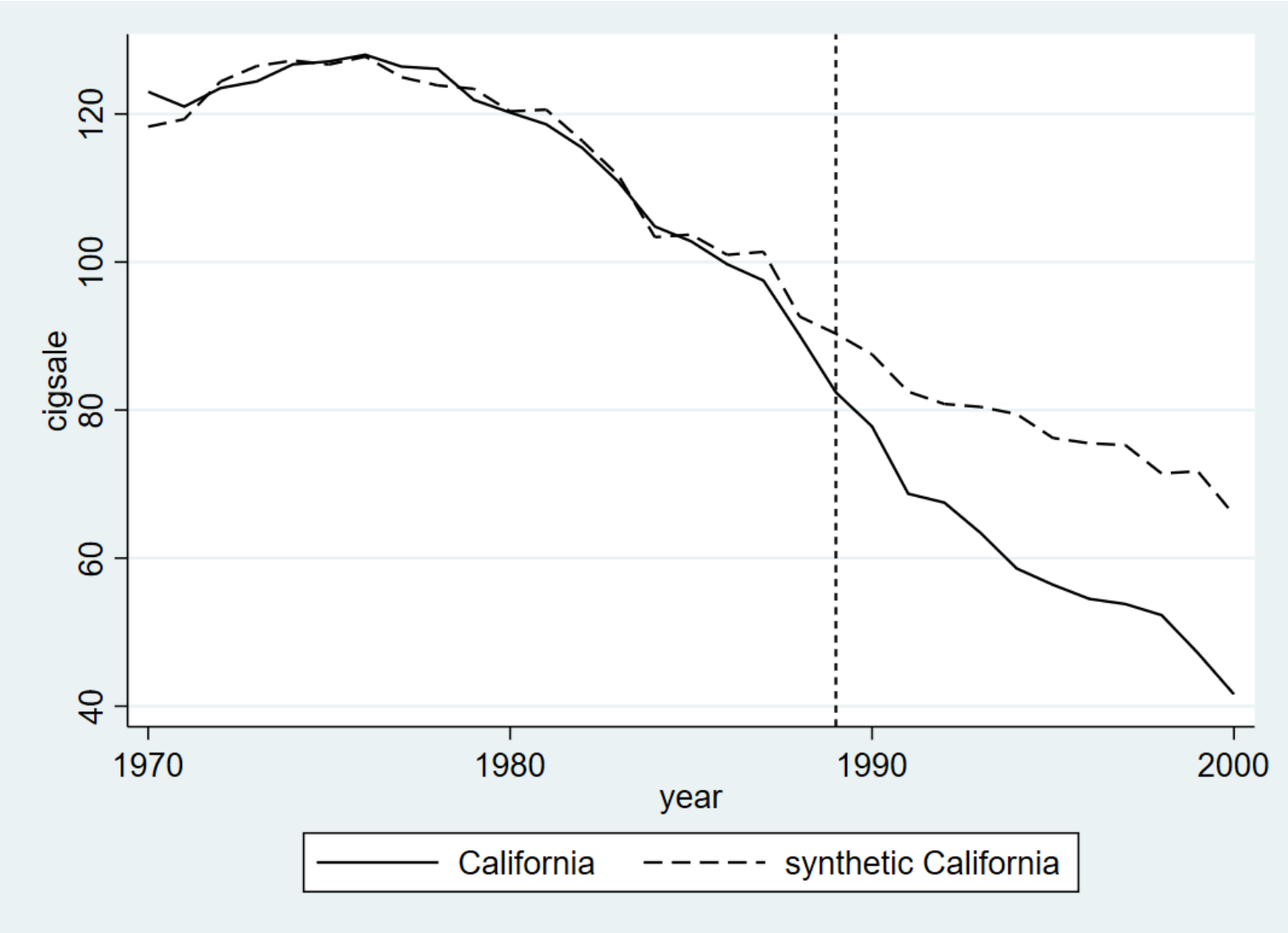
Statistical Inference of SC

- The steps of **Permutation Methods** for inference
  - **Step 1:** Estimate a "placebo" treatment effect for each unit in the control group ("donor pool") using SC methods
  - **Step 2:** Calaulate an empirical p-value for the estimated effect on the treated unit at time t

$$p_{1t} = \frac{\sum_{i=2}^{J+1} \mathbf{1}\{\hat{\alpha}_{it} \geq \hat{\alpha}_{1t}\}}{J}$$

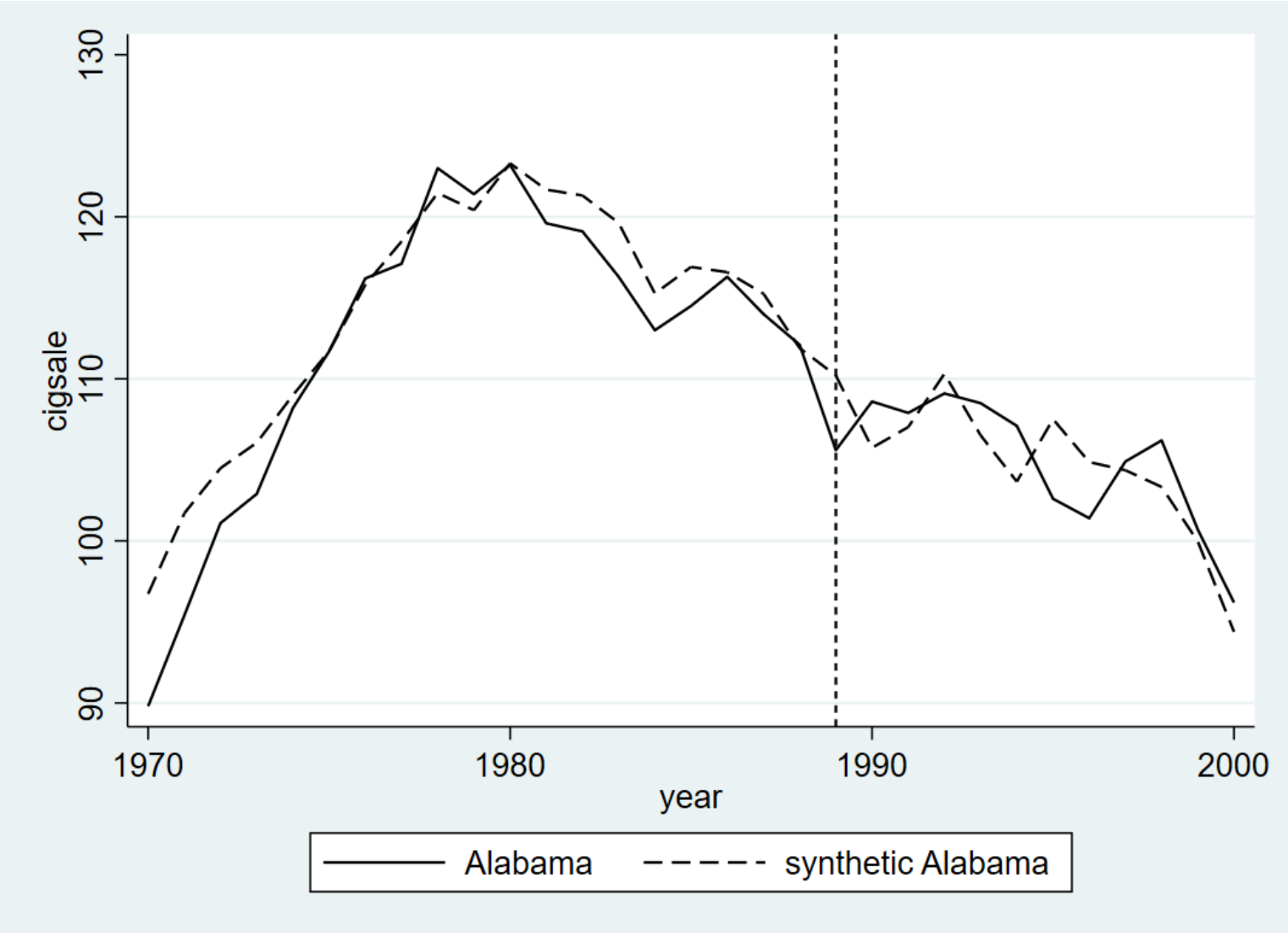
SC Analysis for Treated State

California



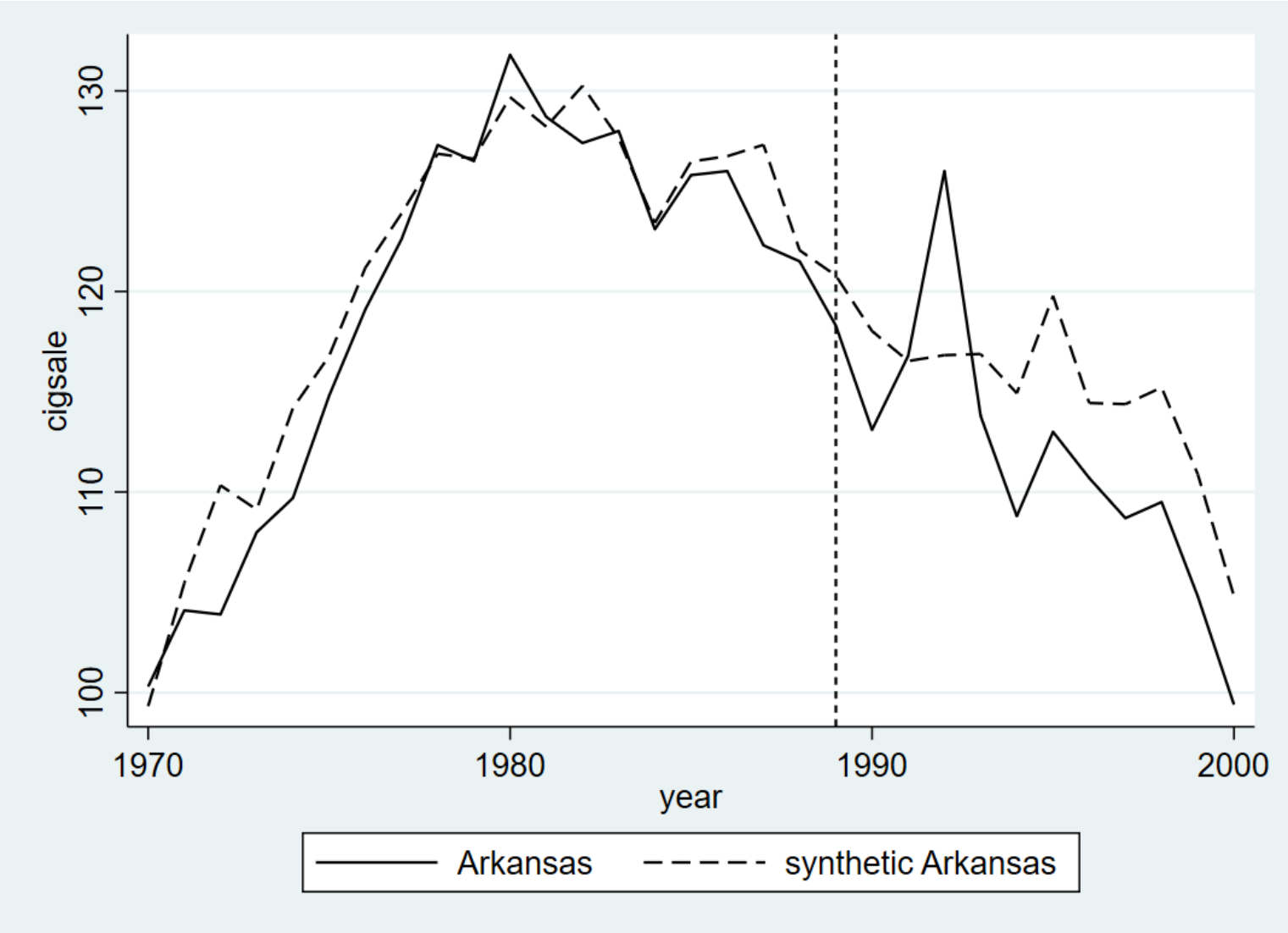
SC Analysis for Non-treated States

Alabama



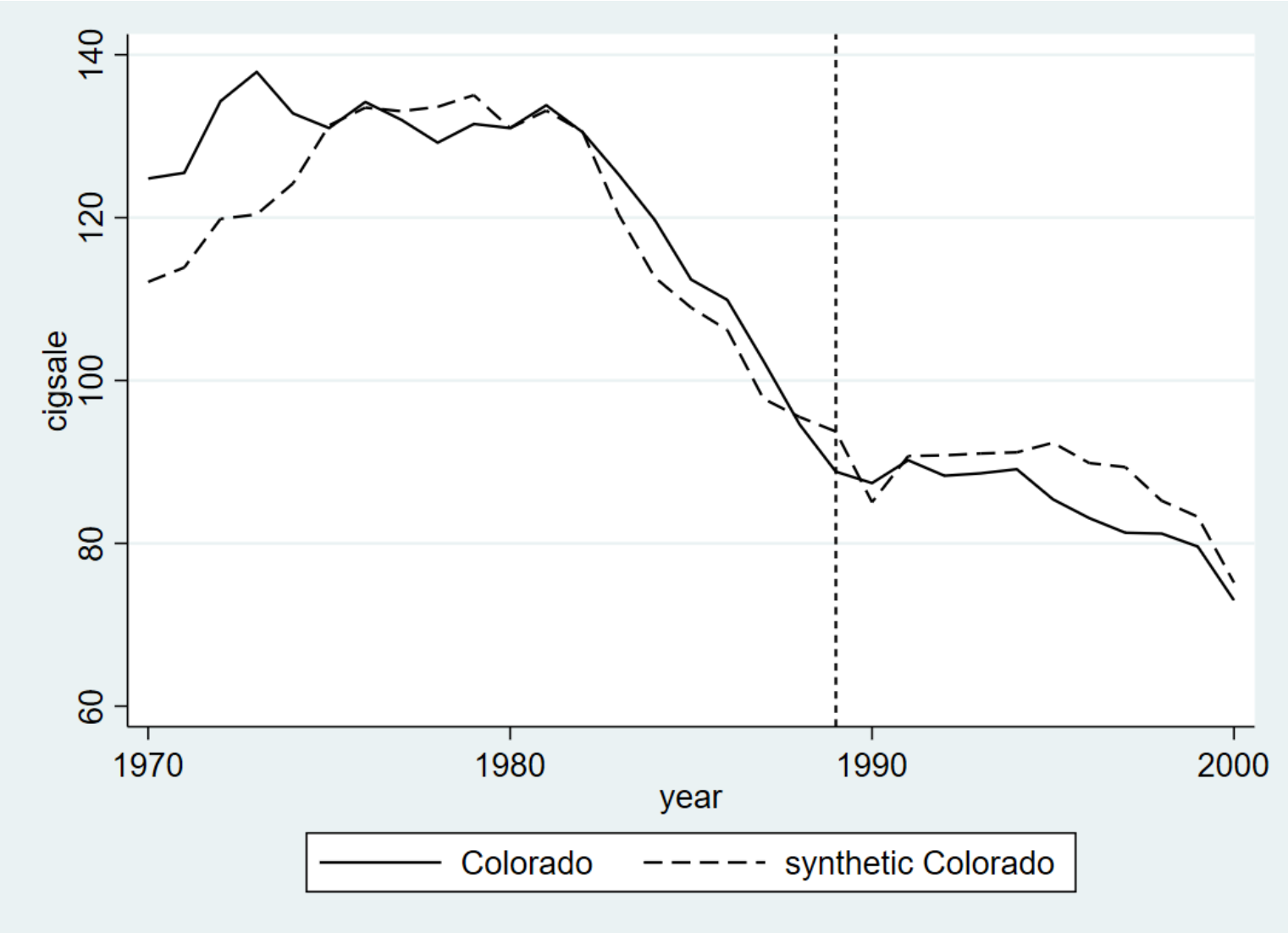
SC Analysis for Non-treated States

Arkansas



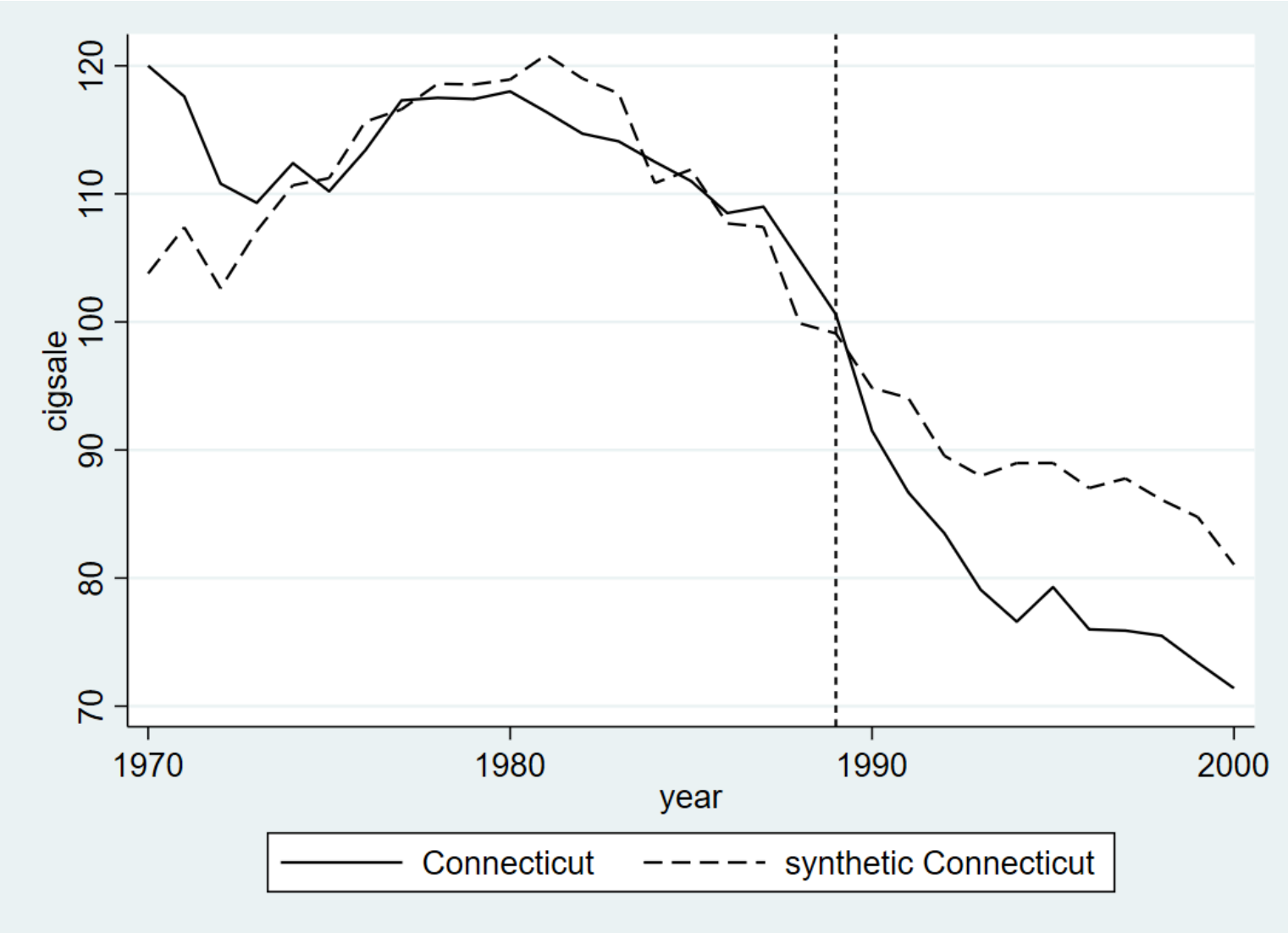
SC Analysis for Non-treated States

Colorado



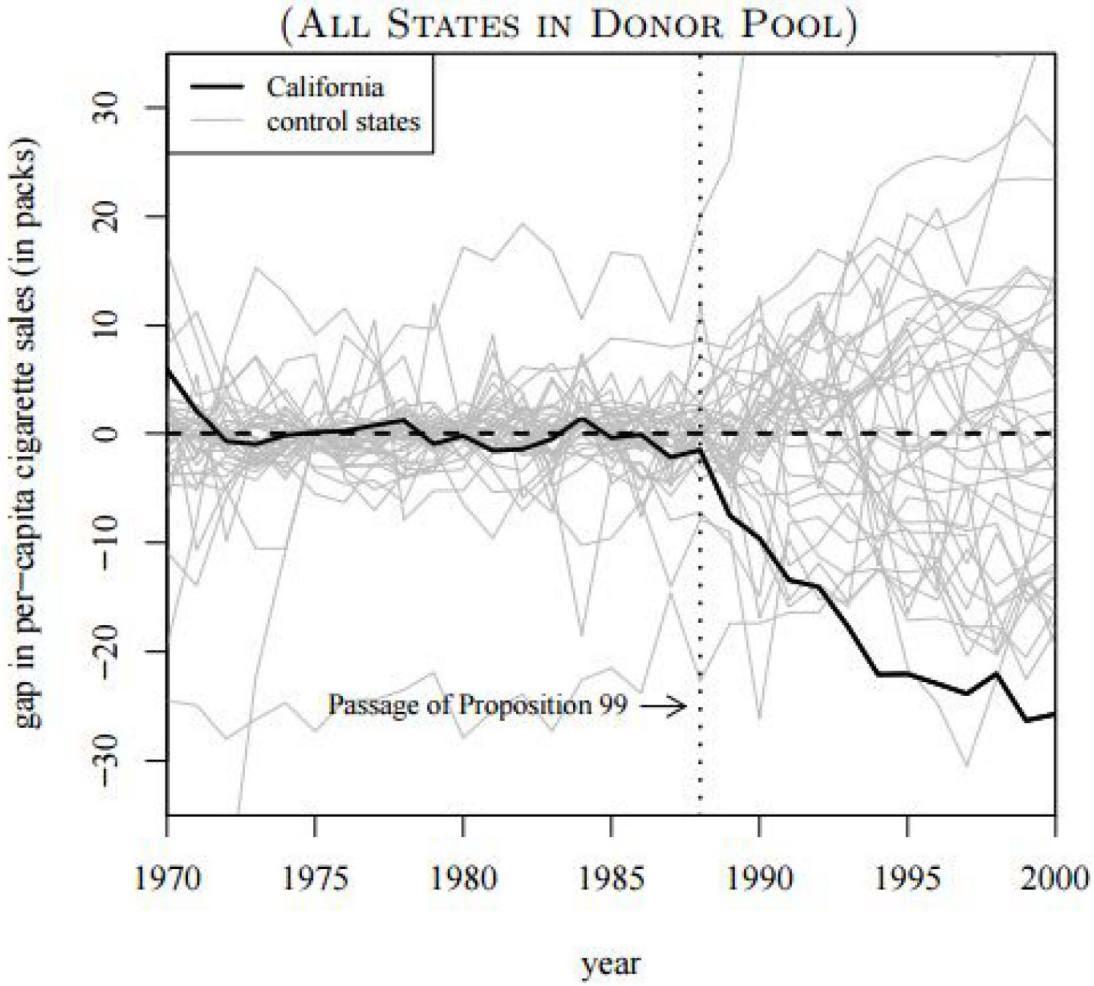
SC Analysis for Non-treated States

Connecticut



SC Estimates for California vs. Non-treated States

Use All Placebo Estimates



Statistical Inference of SC

- The placebo effects may be quite large if those units where NOT matched well in the pre-treatment period
- For example, when the SC fit is bad, we may get erroneous inferences
  - Look at bottom gray line in the above figure
- This would cause p-values to be too conservative
  - Too high p-value

Statistical Inference of SC

- To control for this, one may want to **adjust $\hat{\alpha}_{it}$ for the quality of the pre-treatment matches**
- **Method 1:**
  - Removing observations with too big Root Mean Squared Predictive Error (RMSPE) during the pre-treatment period

$$RMSPE_i^{pre} = \sqrt{\frac{\sum_{t=1}^{T_0}(Y_{it} - Y_{it}^{SC})^2}{T_0}}$$

SC Estimates for CA vs. Placebo Treatments

Use High-Quality Placebo Estimates

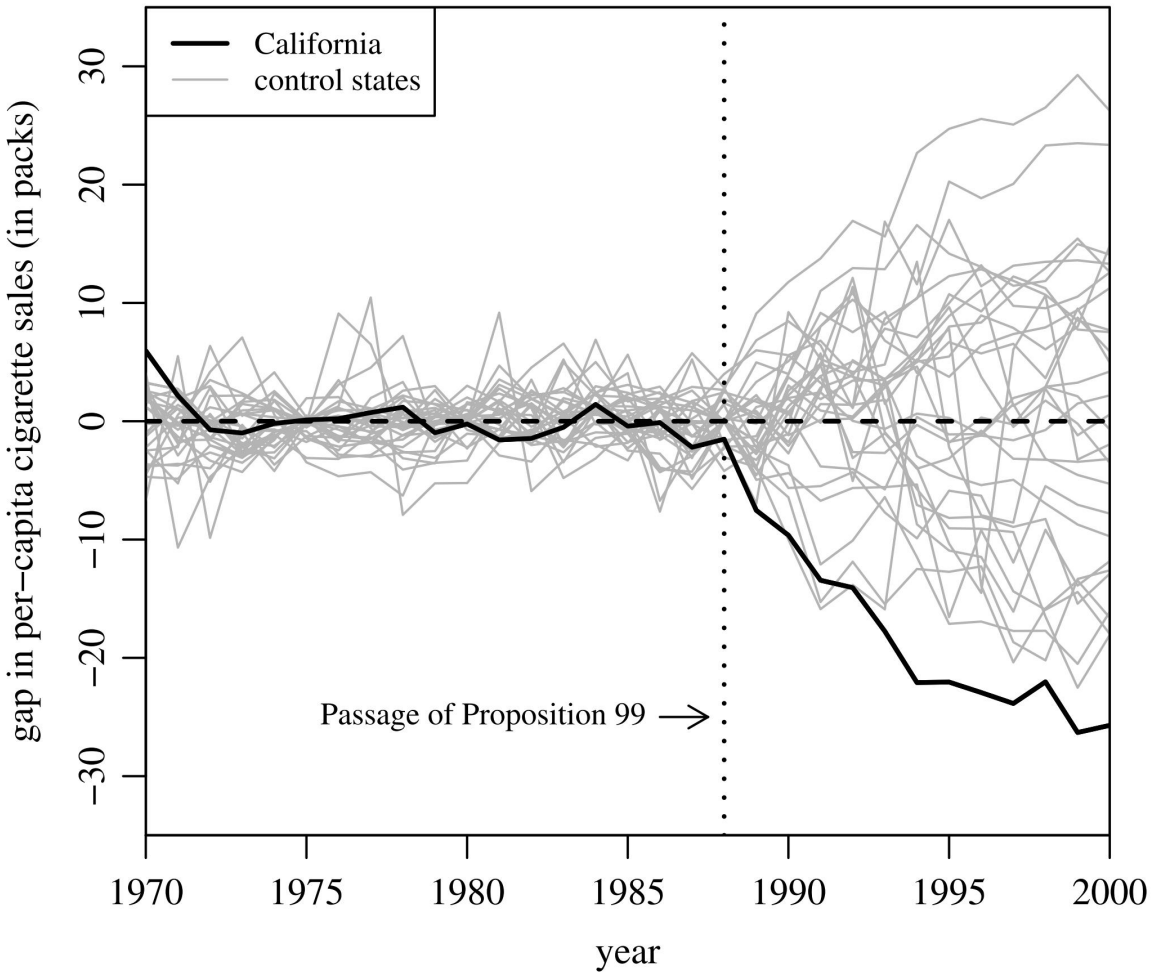


Figure 6. Per-capita cigarette sales gaps in California and placebo gaps in 29 control states (discards states with pre-Proposition 99 MSPE five times higher than California's).

- **Method 2:**
  - We can also calculate p-value by taking pre-treatment matching quality into account:

$$p_{1t}^{std} = \frac{\sum_{i=2}^{J+1} \mathbf{1}\{\frac{\hat{\alpha}_{it}}{RMSPE_i^{pre}} \geq \frac{\hat{\alpha}_{1t}}{RMSPE_1^{pre}}\}}{J}$$

$$RMSPE_i^{pre} = \sqrt{\frac{\sum_{t=1}^{T_0}(Y_{it} - Y_{it}^{SC})^2}{T_0}}$$

Synthetic Control Method: STATA Example

STATA Example: Chen (2022)

The Effect of US-China Trade War on Export

Hsin-Yi Chen (2022), "**The Effect of US-China Trade War on Export: Evidence from China and Taiwan**", Master Thesis

- She wants to examine the effect trade war on China's and Taiwan's export to US

STATA Example: Chen (2022)

STATA Implementation

- See SCM_trade.do
- Use Comtrade_US_All_Countries_sa_used_pwt_ch.dta
- Install the following ado files:
  - synth.ado
  - synth_runner.ado

STATA Example: Chen (2022)

STATA Implementation

- To implement SC method in STATA, we can use two STATA commands:
- `synth`:
  - It is only for **single treated unit** case
  - It can compute optimal weights and covariate balance
- `synth_runner`:
  - It can be used for **single/multiple treated units** case and **do statistical inference** automatically
  - It is designed to accompany `synth` but not to supersede it

STATA Example: Chen (2022)

STATA Implementation

- We can use the following commands to install `synth` and `synth_runner`

```
net from "http://web.stanford.edu/~jhain/Synth/"
net install synth, all replace force
```

```
net install synth_runner, from(https://raw.github
    .com/bquistorff/synth_runner/master/)replace}
```

STATA Example: Chen (2022)

Step 1: Setup Panel Data

```
1 use $wdata/
      Comtrade_US_All_Countries_sa_used_pwt_ch,
      replace
2 tsset country period
```

- Before using SC estimation, dataset must be declared as a (balanced) panel dataset using **tsset** *panelvar timevar*
- **tsset**: Tell STATA which variable is a panel variable and time variable
- We will use this information to define unit number (ID) and time period

```
tsset country period
```

STATA Example: Chen (2022)

Step 2: SC Estimation

- Syntax:

```
synth depvar predictorvars, trunit(#) trperiod(#)
```

- **depvar**: the outcome variable (Y)
- **predictorvars**: the list of predictor variables (Z)
- By default, all predictor variables are **averaged over the entire pre-treatment period**

STATA Example: Chen (2022)

Step 2: SC Estimation

- Syntax:

```
synth depvar predictorvars , trunit(#) trperiod
    (#)
```

- **trunit()**: the unit number (ID) of the unit affected by the treatment
- **trperiod()**: the time period when the treatment occurred
- The above unit number (ID) and time period are specified in **tsset**

STATA Example: Chen (2022)

Step 2: SC Estimation

```
synth TradeValue_growth gdp_growth(2(1)30)
    ex_growth im_growth
TradeValue_growth(2) TradeValue_growth(3)
    TradeValue_growth(4) TradeValue_growth(5)
    TradeValue_growth(6),trunit(34) trperiod(35)
```

- The unit affected by the intervention is unit number 34 (China) in the period 35 (2018 Q3)
- **gdp_growth(2(1)30)**: the value of gdp growth averaged over the periods 2,3,...,30 is entered as a predictor
- **ex_growth** and **im_growth**: since no variable specific period is provided, the value of these variables are averaged either over the entire pre-treatment period
- **TradeValue_growth(4)**: the value of the variable **TradeValue_growt** in the period 4 is entered as a predictor

STATA Example: Chen (2022)

Results: Unit Weights for SC

- Top 3 countries that consist synthetic China
  - Cambodia: 17.7%
  - Slovenia: 10.2%
  - Australia: 9.4%

STATA Example: Chen (2022)

Results: Predictor Means

Predictor Balance:

	Treated	Synthetic
gdp_growth	119.1732	112.2619
cons_growth	107.1192	101.62
inv_growth	96.09201	99.32548
gov_growth	100.1331	107.0138
ex_growth	107.6551	105.4111
im_growth	111.6944	108.6153
pop_growth	102.0774	105.2777
TradeValue_growth(2)	108.8959	107.8139
TradeValue_growth(3)	114.401	114.2687
TradeValue_growth(4)	117.4691	117.2843
TradeValue_growth(5)	118.9544	118.7416

STATA Example: Chen (2022)

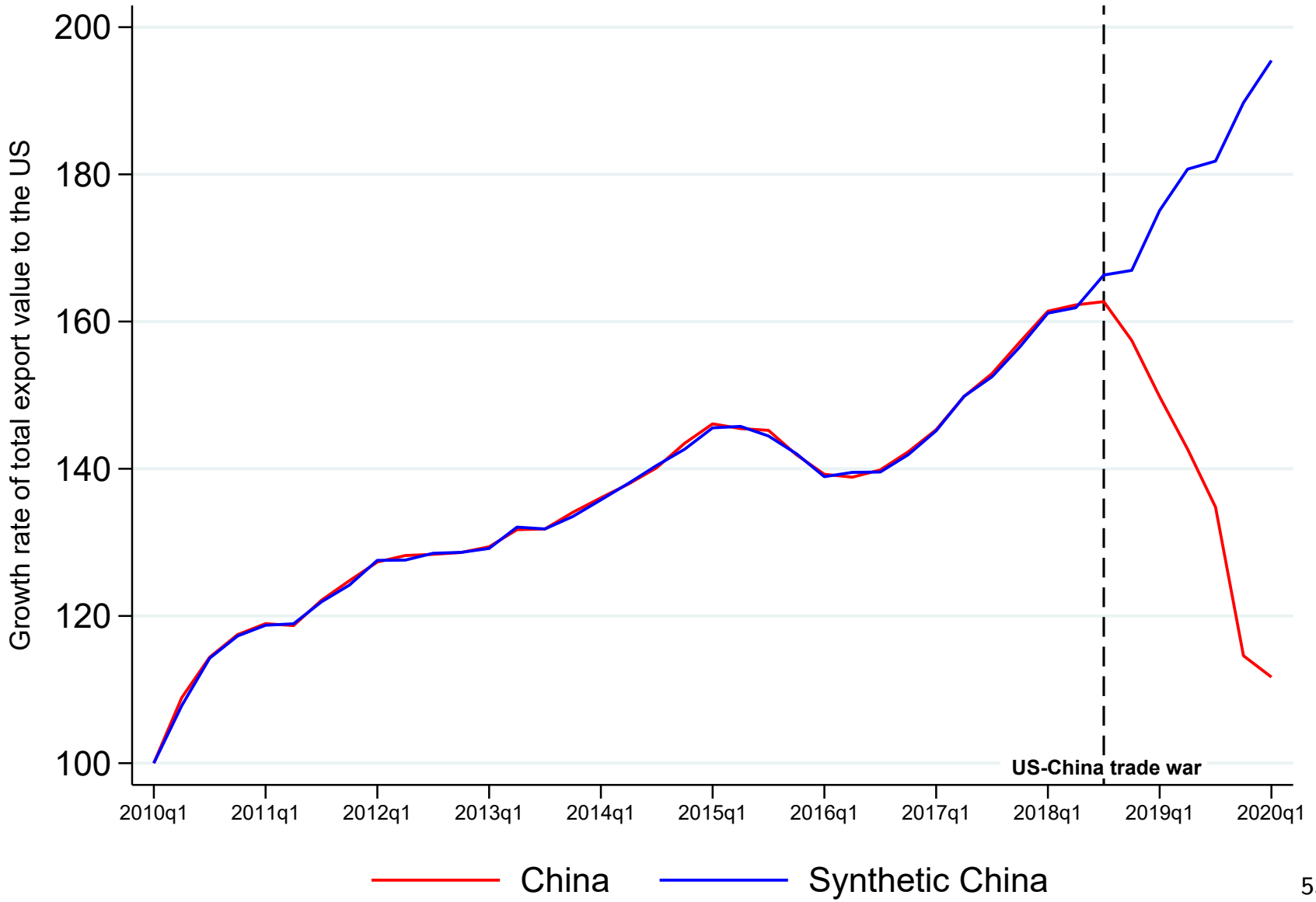
Step 3: SC Graphs

```
synth TradeValue_growth gdp_growth(2(1)30)
    ex_growth im_growth
TradeValue_growth(2) TradeValue_growth(3)
    TradeValue_growth(4) TradeValue_growth(5)
    TradeValue_growth(6),trunit(34) trperiod(35)
    figure
```

- **figure**: produces a line plot with outcome trends for the treated unit and the synthetic control unit

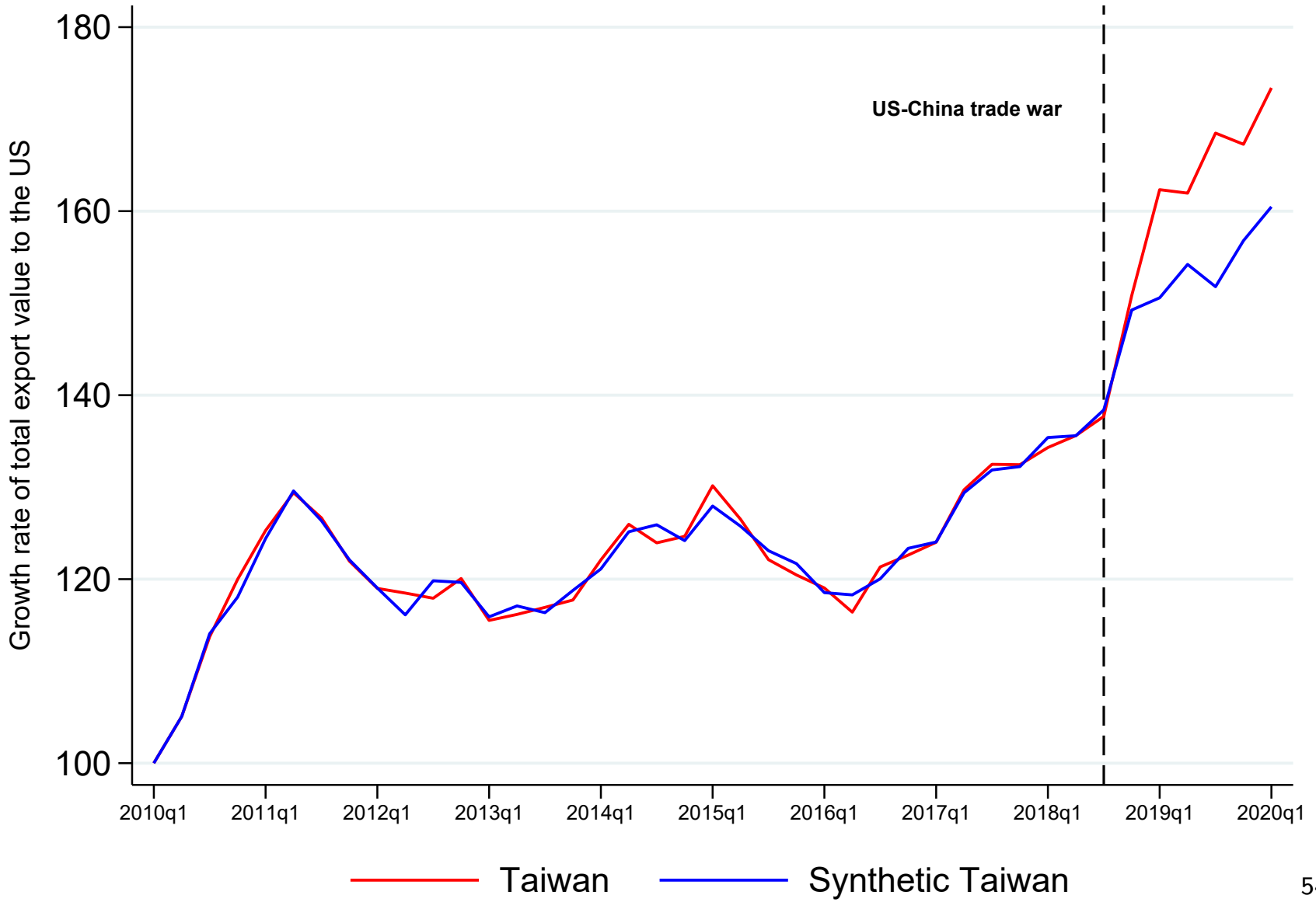
STATA Example: Chen (2022)

Results: SC Graphs



STATA Example: Chen (2022)

Results: SC Graphs



STATA Example: Chen (2022)

Step 4: Other Settings

```
synth TradeValue_growth gdp_growth(2(1)30)
    ex_growth im_growth
TradeValue_growth(2) TradeValue_growth(3)
    TradeValue_growth(4) TradeValue_growth(5)
    TradeValue_growth(6),trunit(34) trperiod(35)
    nested allopt counit(4(1)30) keep("sc_results
    ") replace
```

- **nested**: produces a better fit at the expense of additional computing time
- **allopt**: It can improve the fit even further and requires yet more computing time
- **counit(numlist)**: a list of unit numbers for the control units as given in the panel id variable specified in **tsset**
  - **counit(4(1)30)**: Use unit number 4 – 30 as non-treated units

STATA Example: Chen (2022)

Step 4: Other Settings

```
synth TradeValue_growth gdp_growth(2(1)30)
    ex_growth im_growth
TradeValue_growth(2) TradeValue_growth(4)
    TradeValue_growth(6) TradeValue_growth(8)
    TradeValue_growth(34),trunit(34) trperiod(35)
     nested allopt counit(4(1)30) xperiod(1980(1)
    1988) keep("sc_results") replace
```

- **keep("sc_results")**: Save the SC results to a file
- It contains the following results
  - Estimated SC units
  - Weights for untreated units
- **replace**: Replace the previous file

STATA Example: Chen (2022)

Step 5: Statistical Inference

- Syntax:

```
synth_runner depvar predictorvars, [trunit(#)
    trperiod(#) keep(file) replace pvals1s
    pre_limit_mult(real>=1)]
```

- **pvals1s**: outputs one-sided p-values instead of the two-sided p-values
- **keep(filename)**: save a dataset with the SC results
- **replace**: replace previous files that you have save
- We can merge this dataset which has SC results with the initial dataset and get treatment effect graphs

STATA Example: Chen (2022)

Step 5: Statistical Inference

- Syntax:

```
synth_runner depvar predictorvars, [trunit(#)
    trperiod(#) d(varname) keep(file) pvals1s
    pre_limit_mult(real>=1)]
```

- **pre_limit_mult(real>=1)**: it will exclude placebo effects that has too big Root Mean Squared Predictive Error (RMSPE) in the pre-treatment period

STATA Example: Chen (2022)

Step 5: Statistical Inference

```
synth_runner TradeValue_growth gdp_growth(2(1)30)
    ex_growth im_growth
TradeValue_growth(2) TradeValue_growth(3)
    TradeValue_growth(4) TradeValue_growth(5)
    TradeValue_growth(6),trunit(34) trperiod(35)
    pvals1s pre_limit_mult(5)
```

- Option **pre_limit_mult(5)**
  - Exclude non-treated states with 5 times of RMSPE

STATA Example: Chen (2022)

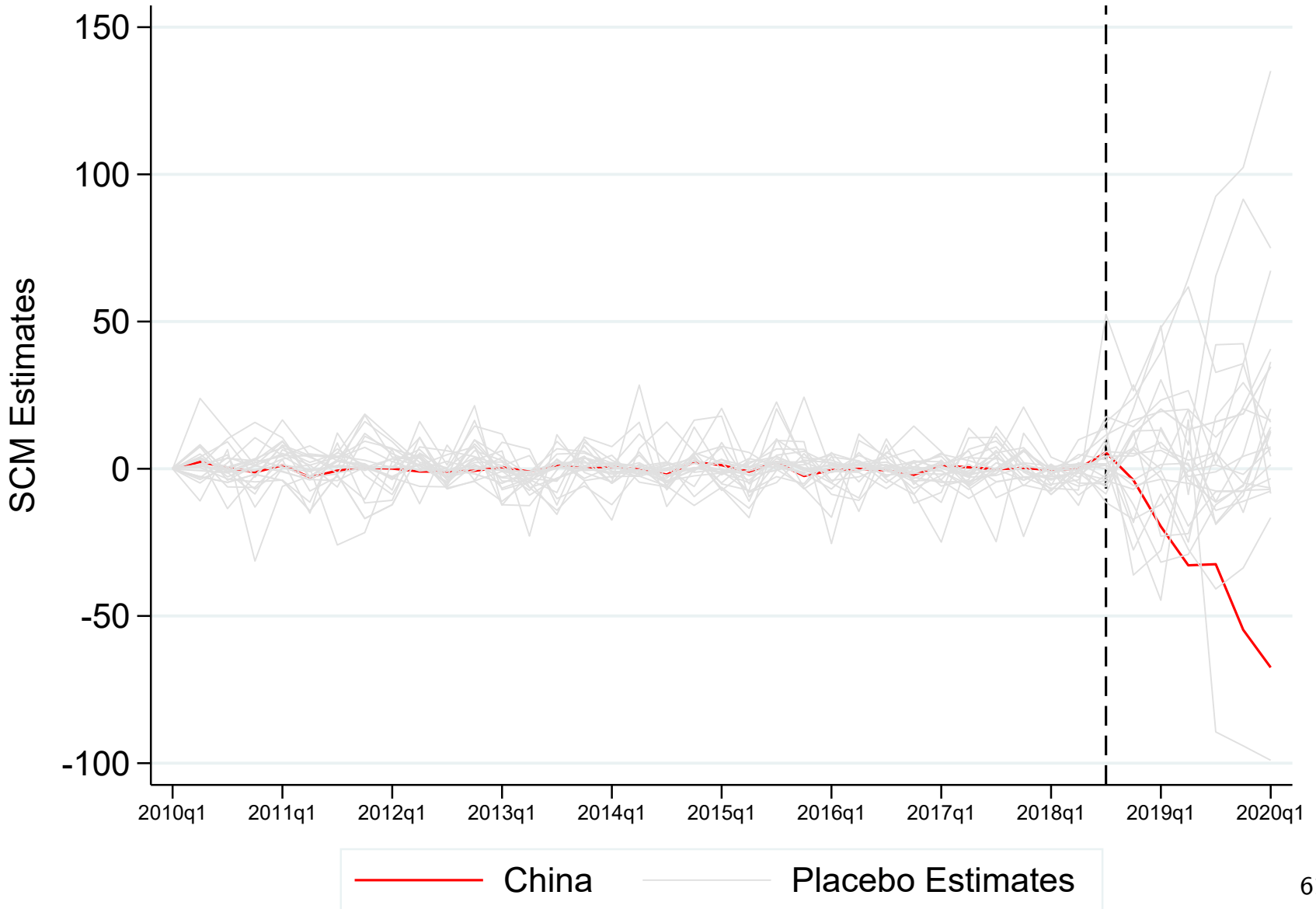
Step 5: Statistical Inference

Post-treatment results: Effects, p-values, standardized p-values

	estimates	pvals	pvals_std
c1	-3.618968	.8382353	.0147059
c2	-9.505467	.6911765	0
c3	-25.31065	.4852941	0
c4	-38.04106	.3529412	0
c5	-47.00758	.2647059	0
c6	-75.13612	.1911765	0
c7	-83.7629	.2205882	0

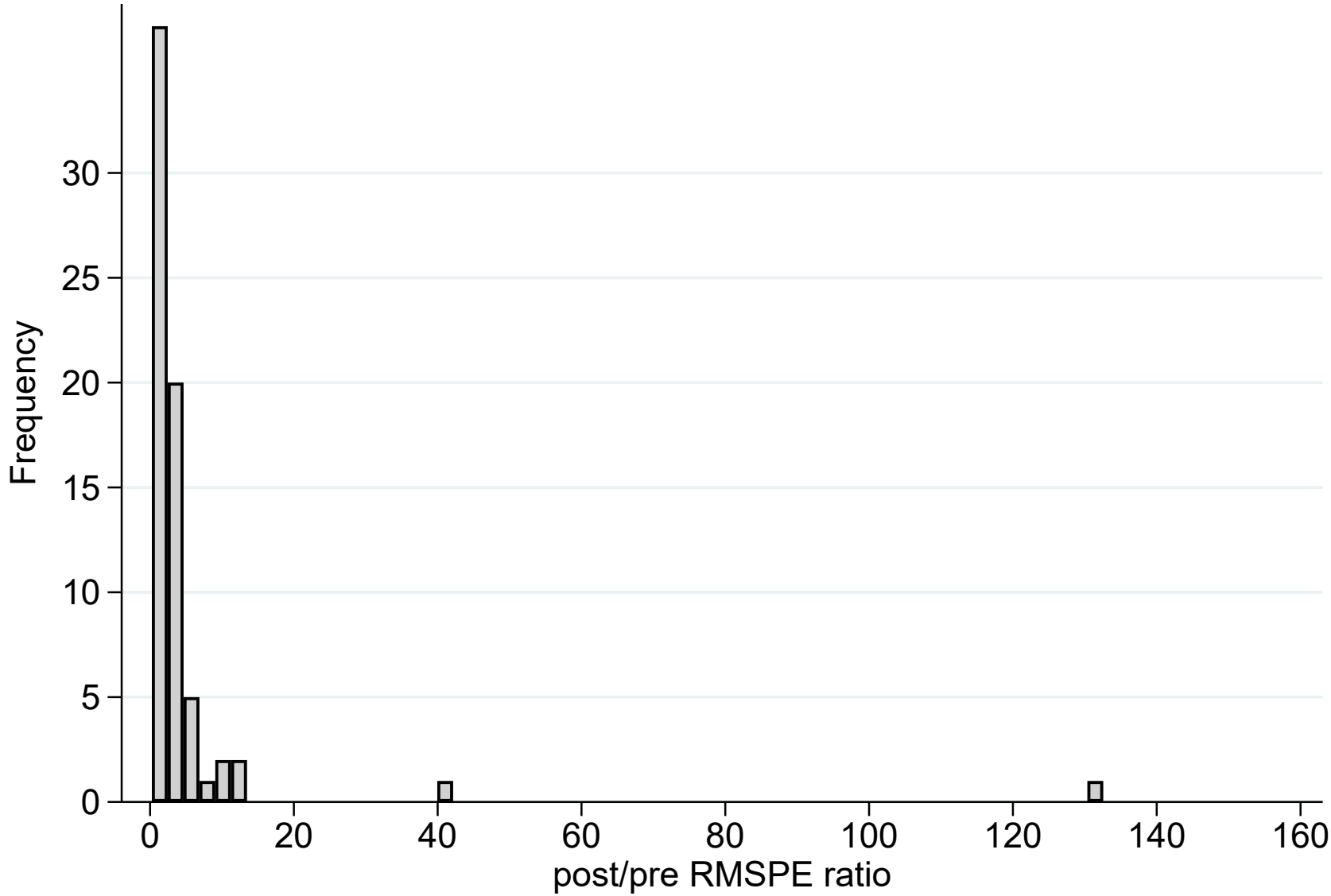
STATA Example: Chen (2022)

Step 5: Statistical Inference



STATA Example: Chen (2022)

Step 5: Statistical Inference



Synthetic Control Method: Multiple Treated Units

Extending the Model to Multiple Treated Units

- If we have multiple treated unit, the procedure of SC estimation is the following:
- **Step 1:** Estimate a separate α_i vector for each treated i using only non-treated units when forming W_i^*
- **Step 2:** Calculate the ATT by"integrating" over G treated units

$$\bar{\alpha} = \{\bar{\alpha}_{T_0+1}, ..., \bar{\alpha}_T\} = \frac{1}{G}\sum_{i=1}^{G}\{\hat{\alpha}_{iT_0+1}, ..., \hat{\alpha}_{iT}\}$$

  - G is number of treated units
  - Therefore, similar to the single treated case, we have dynamic estimates of the ATT

Statistical Inference: Multiple Treated Units

- Several recent methods have been developed for inference with **more than one treated units and at possibly different treatment timings**
- These are also **permutation methods**

Example

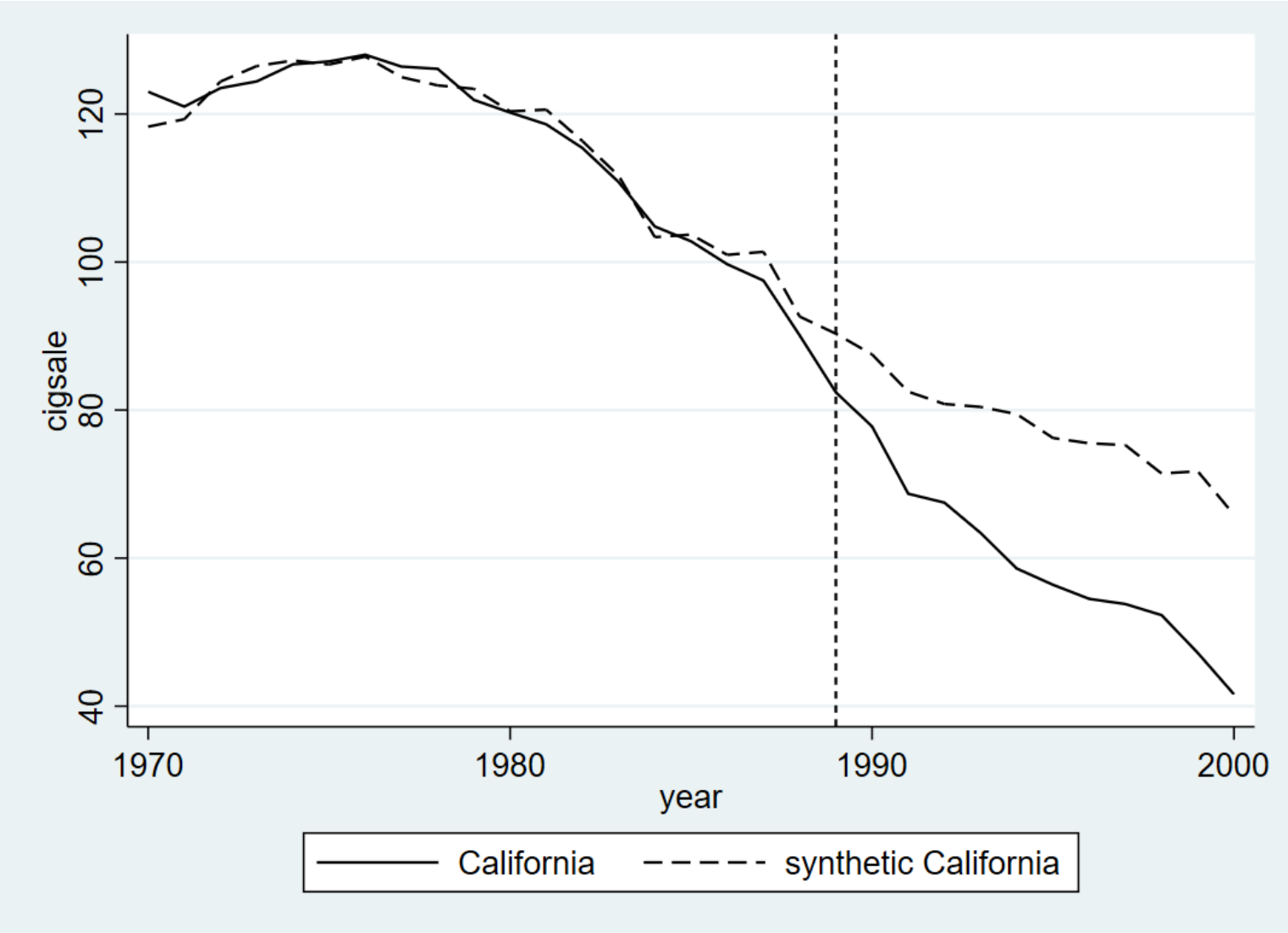
Statistical Inference: Multiple Treated Units

- Suppose Pennsylvania and California pass anti-tobacco law in 1988 and 1987, respectively
  - You want to evaluate its **average treatment effect on treated**
- **Step 1:** Compute the average treatment effect for treated units

$$\bar{\alpha} = \frac{1}{2} \sum_{g=1}^{2} \hat{\alpha}_g$$

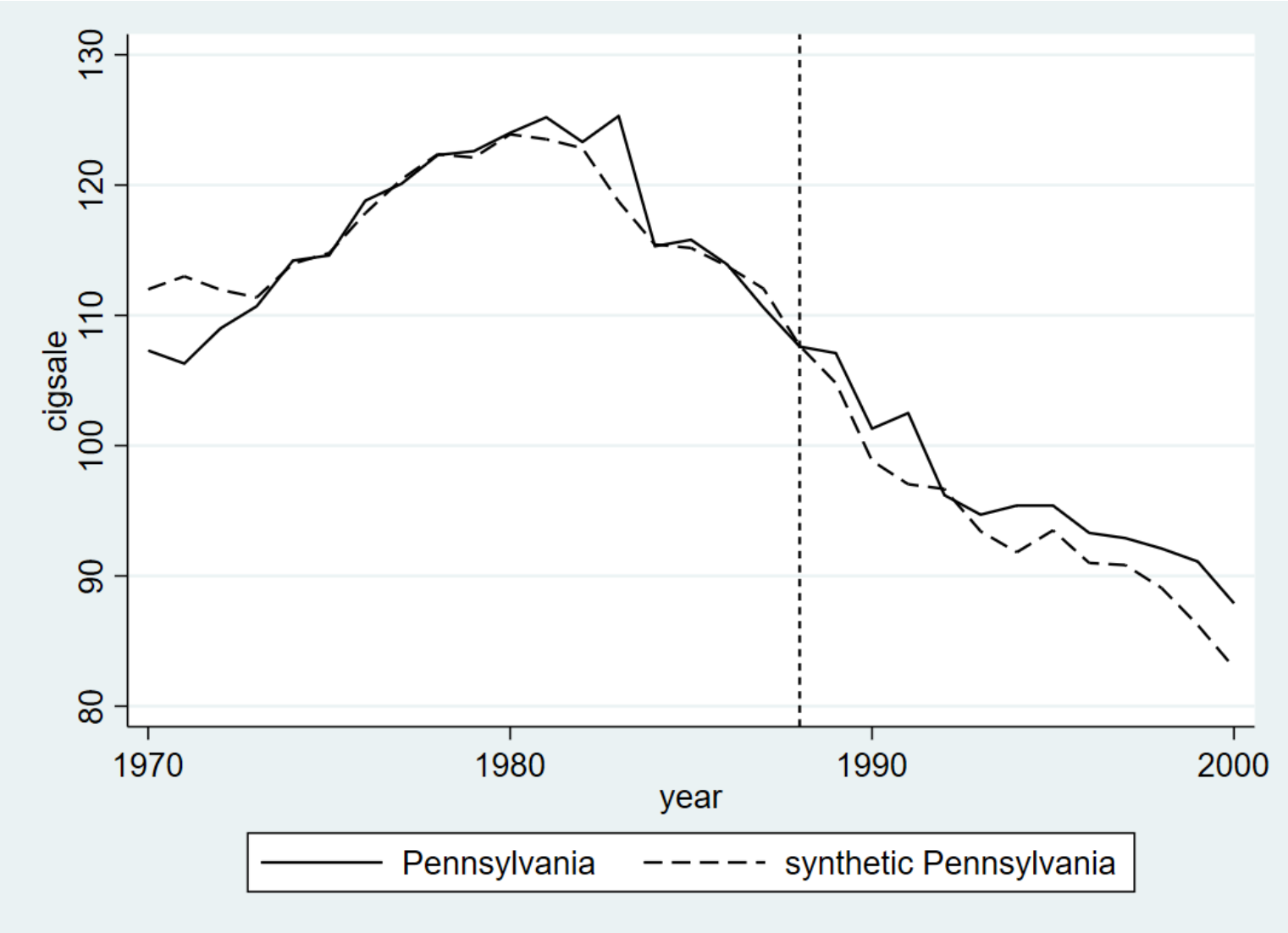
Estimate $\bar{\alpha}$

SC Analysis for Treated States: California



Estimate $\bar{\alpha}$

SC Analysis for Treated States: Pennsylvania



Example

Statistical Inference: Multiple Treated Units

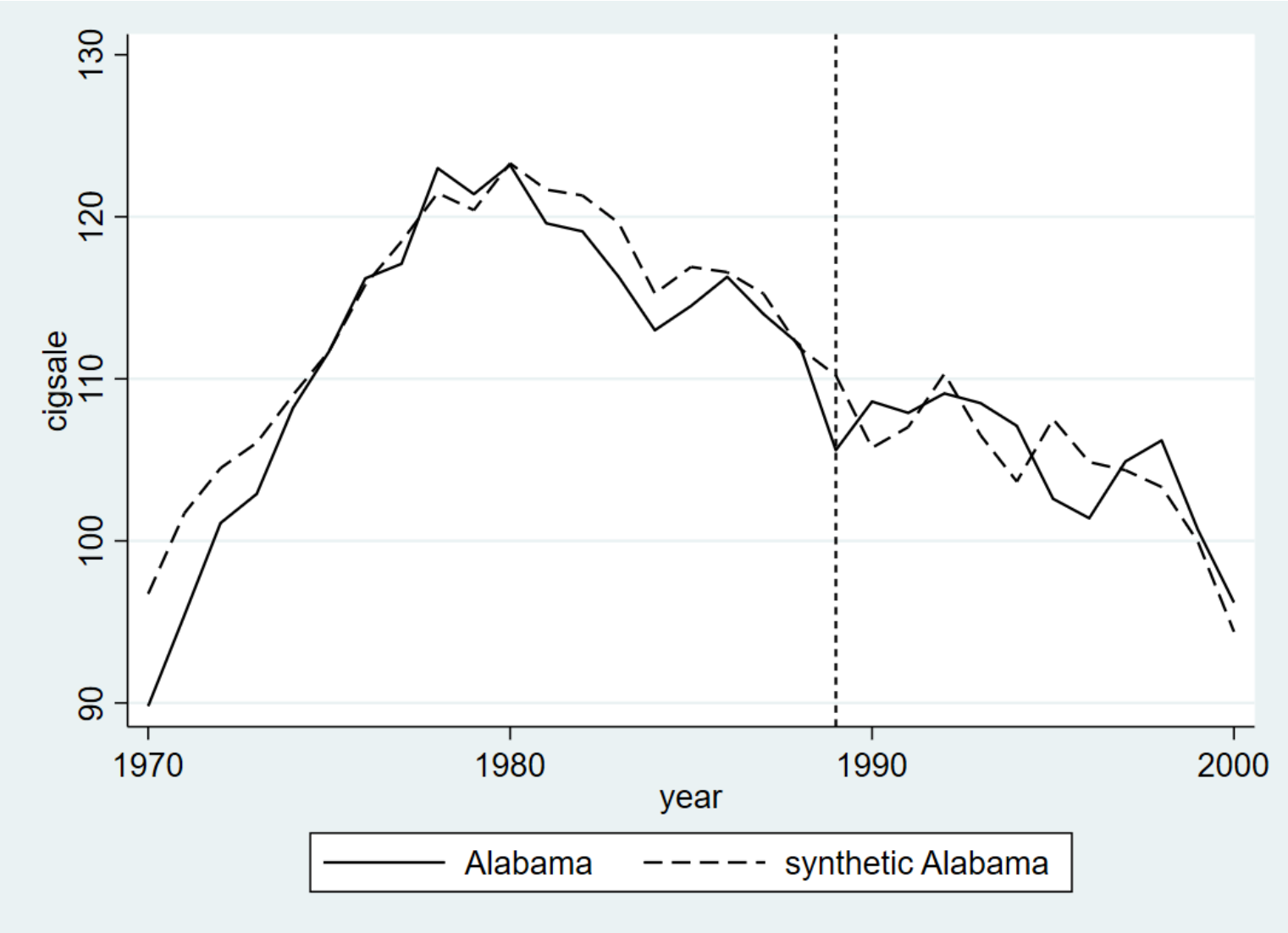
- **Step 2:** Averaging over placebo effect for treated units (Pennsylvania and California)

$$\bar{\alpha}^{PL(i)} = \frac{1}{2}\sum_{g=1}^{2} \hat{\alpha}_g^{PL}$$

- Let i index one possible average placebo effect for treated units
- There are $N_{PL} = \prod_{g=1}^{2} J_g$ such possible averages
  - J_1 is number of placebo effects for Pennsylvania: 37 states
  - J_2 is number of placebo effects for California: 37 states
  - So that $N_{PL} = 37 \times 37 = 1369$ possible averages

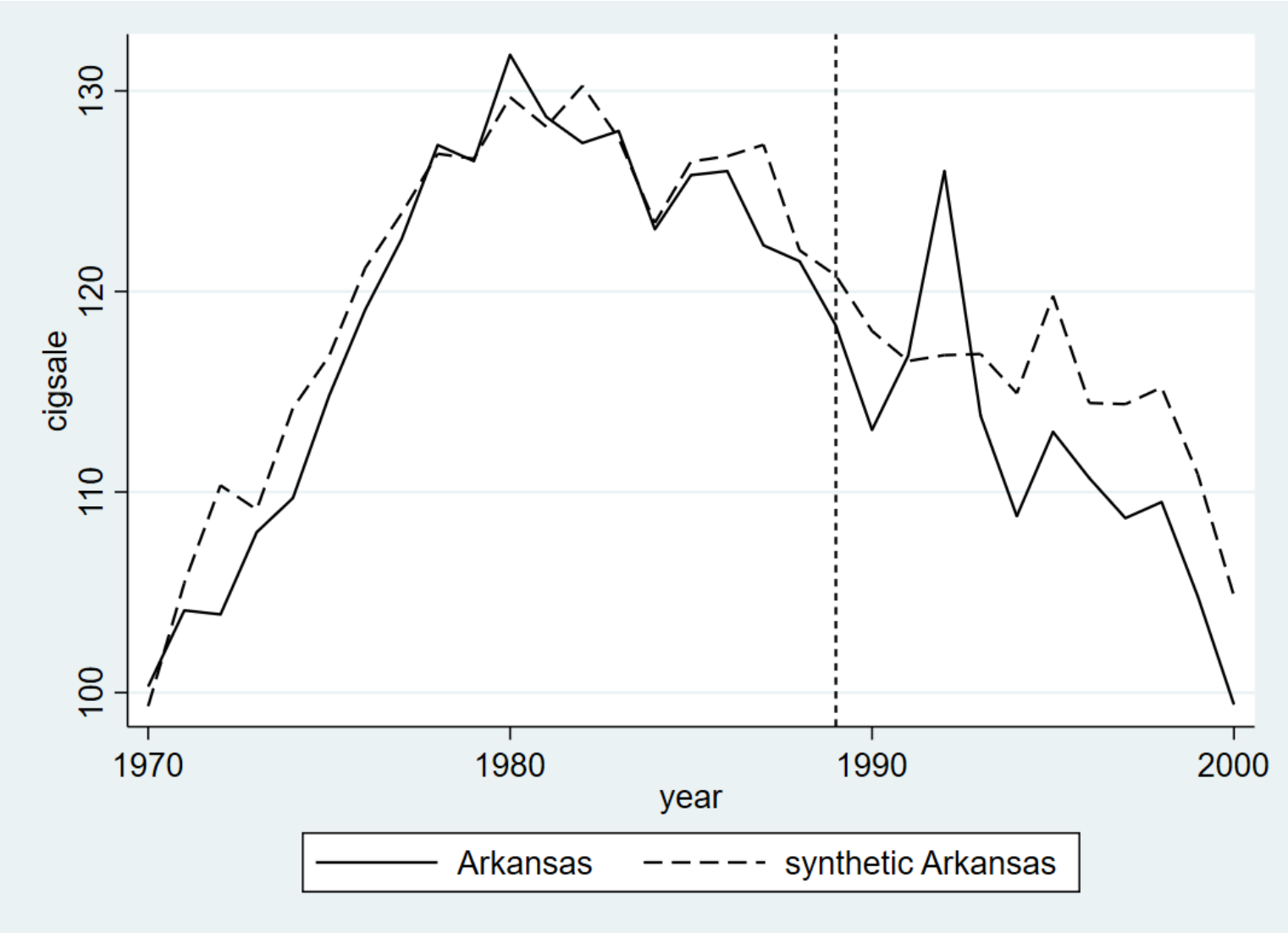
Estimate $\bar{\alpha}^{PL(1)}$

SC Analysis for Pennsylvania's Non-treated States: Alabama



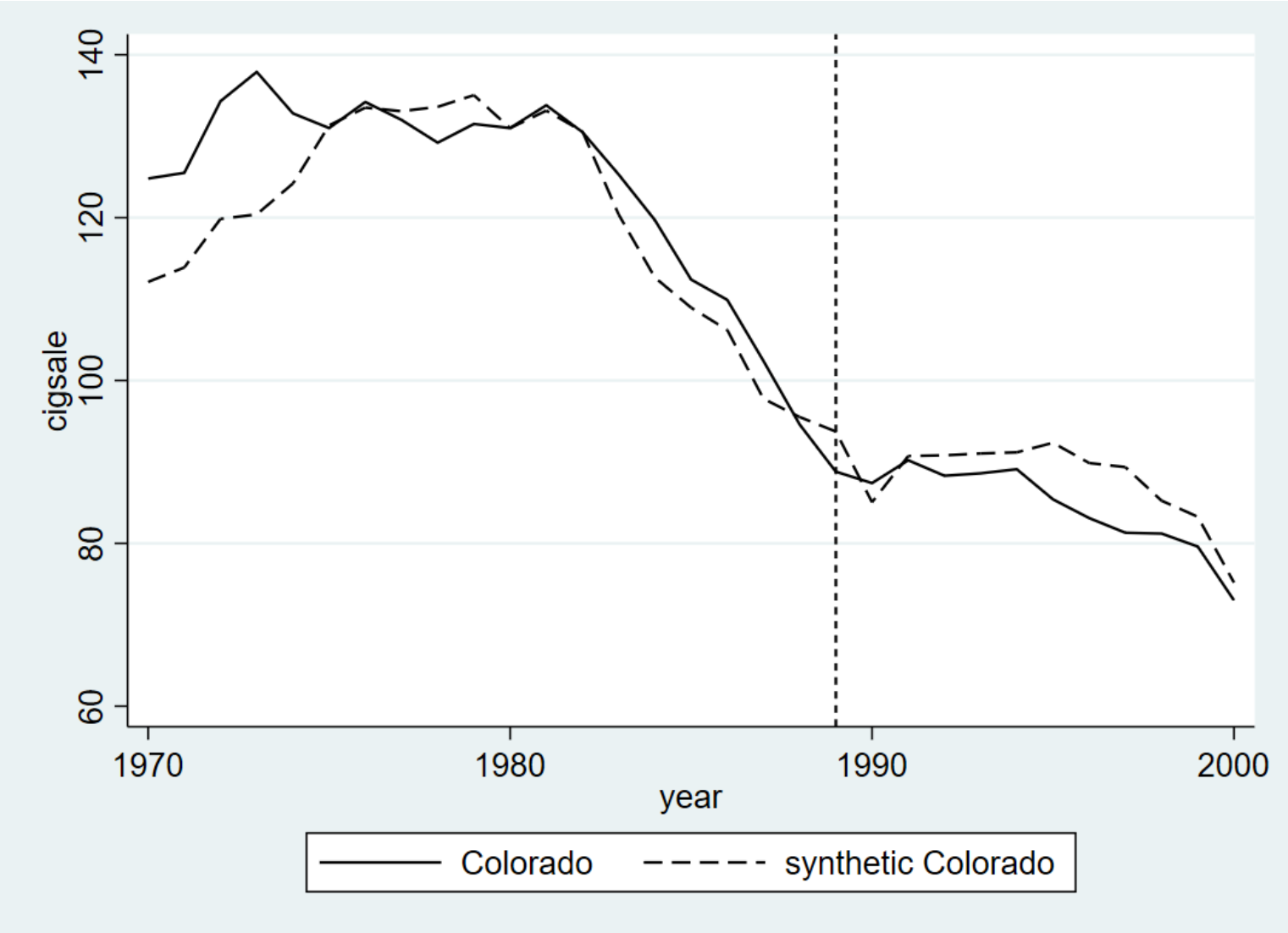
Estimate $\bar{\alpha}^{PL(1)}$

SC Analysis for California's Non-treated States: Arkansas



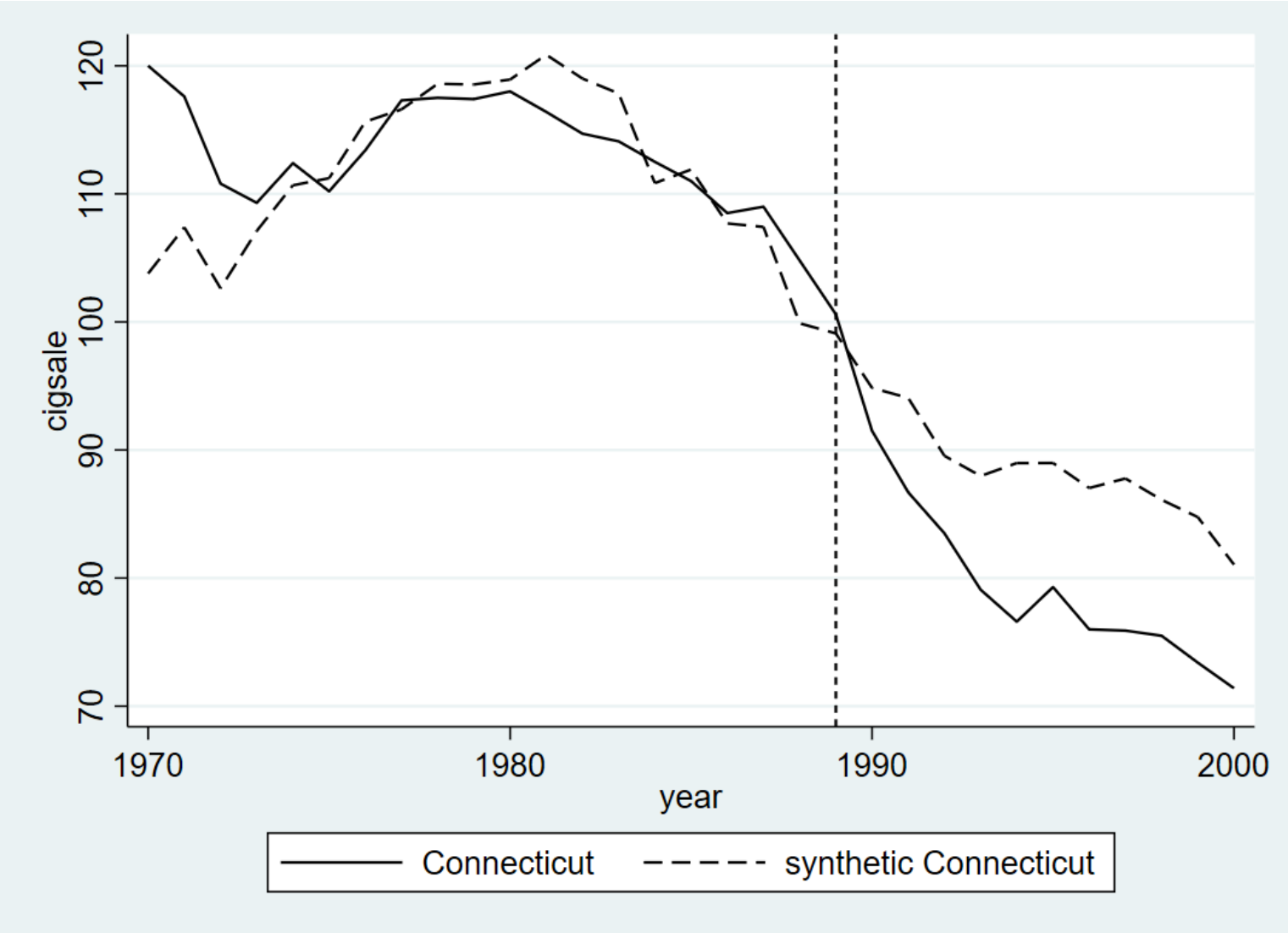
Estimate $\bar{\alpha}^{PL(2)}$

SC Analysis for Pennsylvania's Non-treated States: Colorado



Estimate $\bar{\alpha}^{PL(2)}$

SC Analysis for California's Non-treated States: Connecticut



Example

Statistical Inference: Multiple Treated Units

- **Step 3:** Compute the p-value

$$\text{p-value} = \frac{\sum_{i=1}^{1369} \mathbf{1}\{\bar{\alpha}^{PL(i)} \geq \bar{\alpha}\}}{1369}$$

Statistical Inference: Multiple Treated Units

Cavallo et. al (2013) Method

- Cavallo et. al (2013) suggest the following procedure for statistical inference with multiple treated units
- Let $g = 1, 2..., G$ index treated units and $j = 1, 2...., J$ index non-treated units
  - $G \equiv$ number of treated units
  - $J \equiv$ number of non-treated units
- For a particular treated unit g, one can use SC method to estimate the causal effect $\hat{\alpha}_g$
  - We omit the t subscript as treatment dates may differ across units

Statistical Inference: Multiple Treated Units

Cavallo et. al (2013) Method

- **Step 1:** Compute the average treatment effect for treated units

$$\bar{\alpha} = \frac{1}{G} \sum_{g=1}^{G} \hat{\alpha}_g$$

- For each treatment effect $\hat{\alpha}_g$, we have a corresponding set of placebo effects $\hat{\alpha}_g^{PL}$
- Where each non-treated unit is thought of as entering treatment at the same time as treated unit g
- If two treated units have the same treatment period, then their placebo sets will be the same

Statistical Inference: Multiple Treated Units

Cavallo et. al (2013) Method

- **Step 2:** Averaging over placebo effect for treated units

$$\bar{\alpha}^{PL(i)} = \frac{1}{G}\sum_{g=1}^{G}\hat{\alpha}_g^{PL}$$

- Let i index one possible average placebo effect for treated units
- There are $N_{PL} = \prod_{g=1}^{G} J_g$ such possible averages
  - J_g: number of placebo effects (non-treated units) for treated unit g
  - Thus, $i = 1....N_{PL}$

Statistical Inference: Multiple Treated Units

Cavallo et. al (2013) Method

- **Step 3:** Compute the p-value

$$\text{p-value} = \frac{\sum_{i=1}^{N_{PL}} \mathbf{1}\{\bar{\alpha}^{PL(i)} \geq \bar{\alpha}\}}{N_{PL}}$$

STATA Example: Multiple Treated Units

STATA Implementation

Multiple Treated Units

- See SCM.do
- Use smoking_ca.dta

STATA Implementation

Multiple Treated Units

- Syntax:

```
synth_runner depvar predictorvars, [d(varname)
    pvals1s pre_limit_mult(real>=1)]
```

- The **d(varname)** is a binary variable which is 1 for treated units in treated periods, and 0 everywhere else
- This allows for multiple units to undergo treatment, possibly at different times

STATA Implementation

Multiple Treated Units

- Example:

```
gen byte D = (state==3 & year>=1989) | (state==28
     & year>=1988)
```

- Generate a dummy variable indicate treatment units and post-treatment period
- Treated units: Pennsylvania and California
- Note that no treatment actually happened in Georgia in 1987

STATA Implementation

Multiple Treated Units

- Example:

```
1 synth_runner cigsale beer(1984(1)1987) lnincome
      (1972(1)1987) retprice age15to24, d(D)
      pre_limit_mult(5)
2 effect_graphs
3 pval_graphs
```

- **effect_graphs**: graph average treatment effect
- **pval_graphs**: graph p-value

Synthetic Control Method: Practical Issues

Practical Issues with SC

Computation Time

- The STATA command `synth` implements an optimization procedure to select the optimal W^* for each treated i
- The optimization scheme suffers from a curse of dimensionality
  - Lots of non-treated units increase the number of potential W's
- One way to reduce computation time is to only use non-treated units that are similar to treated units
  - It also improve match quality

Practical Issues with SC

Computation Time

- Many of the inference methods require the use of a large number of placebo treatment groups
  - Single treated problems only require placebo treated **units**, multiple treated problems require placebo **groups**
- Thus, it may be computationally infeasible to implement the Cavallo et. al (2013) method with a large G and J

Practical Issues with SC

Length of Pre-treatment Period

- For the SC estimator to estimate the parameter of interest, the "**pre-treatment period must be large relative to the scale of the transitory shocks**"
  - This is ambiguous, and it is unclear how to test necessary conditions of whether this holds in the data

Suggested Readings

- Chapter 10, Causal Inference: The Mixtape
- Abadie, A. (2019). Using synthetic controls: Feasibility, data requirements, and methodological aspects. Journal of Economic Literature.