Lecture 11:
Synthetic Control, or, Matching 2 of 2

April 2, 2025

Course Administration

1. Quantitative summary feedback posted
2. Workshop is next week
   - post material by Sunday midnight
   - post feedback for classmates before class
   - can agree on later within-group deadline as you like
3. Instructions for presentations posted

Course Administration

1. Quantitative summary feedback posted
2. Workshop is next week
   - post material by Sunday midnight
   - post feedback for classmates before class
   - can agree on later within-group deadline as you like
3. Instructions for presentations posted
4. Instructions for final paper posted – Lec. 9
5. Please come see me about your replication paper
6. Presentations April 16 and 23
7. Paper due April 28, 5 pm
8. Any other issues?

How to Present Things You Want to Compare

Original Table

	Treated	Untreated
Hats	6	5
Coats	8	6

Replication Table

	Treated	Untreated
Hats	6.2	5.1
Coats	8.1	6.1

How to Present Things You Want to Compare

Original Table with Your Work

	Original		Replication	
	Treated	Untreated	Treated	Untreated
Hats	6	5	6.2	5.1
Coats	8	6	8.1	6.1

Even Better

	Original		Replication		Percent Diff.	
	Treated	Untreated	Treated	Untreated	Treated	Untreated
Hats	6	5	6.2	5.1	0.03	0.02
Coats	8	6	8.1	6.1	0.10	0.02

Today

Synthetic control how-to

1. Overview
2. Set-up
3. Goal
4. Convex hull
5. Inference
6. Estimation
7. Example

Today

Synthetic control how-to

1. Overview
2. Set-up
3. Goal
4. Convex hull
5. Inference
6. Estimation
7. Example

Synthetic control example

1. Research question
2. Outcomes
3. Estimation strategy
4. Results
5. What did you think?

Overview of Synthetic Control

When to Use Synthetic Control?

- When we want to know the effect of a policy that happens in one or a few treated places/instances

When to Use Synthetic Control?

- When we want to know the effect of a policy that happens in one or a few treated places/instances
- Why not diff-in-diff?

When to Use Synthetic Control?

- When we want to know the effect of a policy that happens in one or a few treated places/instances
- Why not diff-in-diff?
  - small sample size $\rightarrow$ big standard errors
  - diff-in-diff requires that differences between treated and control are either
    - time-invariant, unit-specific measures or
    - time-varying in the same way for all units

When to Use Synthetic Control?

- When we want to know the effect of a policy that happens in one or a few treated places/instances
- Why not diff-in-diff?
  - small sample size $\rightarrow$ big standard errors
  - diff-in-diff requires that differences between treated and control are either
    - time-invariant, unit-specific measures or
    - time-varying in the same way for all units
- We can weaken these assumptions by making a synthetic control
  - compare treated state to
  - comparison state that is a little of Michigan, a little of Illinois, no Wisconsin and a little Florida

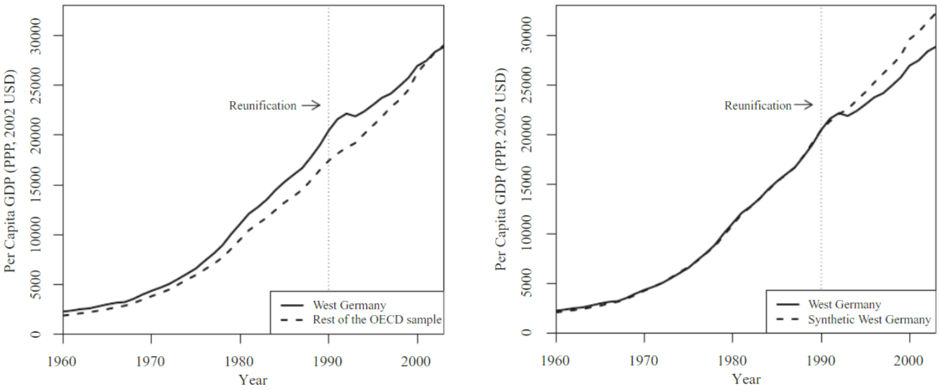
When to Use Synthetic Control?

- When we want to know the effect of a policy that happens in one or a few treated places/instances
- Why not diff-in-diff?
  - small sample size $\rightarrow$ big standard errors
  - diff-in-diff requires that differences between treated and control are either
    - time-invariant, unit-specific measures or
    - time-varying in the same way for all units
- We can weaken these assumptions by making a synthetic control
  - compare treated state to
  - comparison state that is a little of Michigan, a little of Illinois, no Wisconsin and a little Florida
- This doesn't fix the small sample problem, but we use different inference methods

Basic Intuition: Impact of German Reunification

FIGURE 1 Trends in per Capita GDP: West Germany versus Rest of the OECD Sample

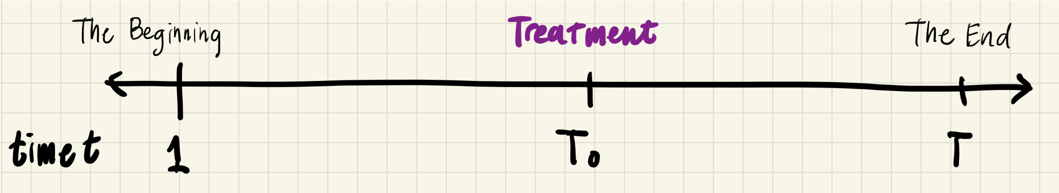
FIGURE 2 Trends in per Capita GDP: West Germany versus Synthetic West Germany



Abadie, Diamond and Hainmueller, "Comparative Politics and the Synthetic Control Method," 2015. [link]

Set-up

Time in Synthetic Control



- t is time, $t \in \{1, ..., T_0, ..., T\}$.
- Treatment occurs after T_0.
- We look for effects starting in $T_0 + 1$
  - there are T_0 pre-intervention periods, $\{1, ..., T_0\}$
  - there are T_1 post-intervention periods, $\{T_0 + 1..., T\}$
  - total $T = T_0 + T_1$

Units in Synthetic Control

- Each unit is denoted i
- There are $J+1$ total units, $i \in \{1..., J+1\}$
- Unit 1 is treated,
- Units $\{2, 3, ..., J+1\}$ are not treated
- $\rightarrow J$ untreated observations
- Call untreated observations the "donor pool"

Conditions for Donor Pool Observations

Conditions for Donor Pool Observations

- Untreated during observation period
  - no policy in the donor pool observations

Conditions for Donor Pool Observations

- Untreated during observation period
  - no policy in the donor pool observations
- Unaffected by treatment
  - no spillover from treated observations

Conditions for Donor Pool Observations

- Untreated during observation period
  - no policy in the donor pool observations
- Unaffected by treatment
  - no spillover from treated observations
- Should have no large, "idiosyncratic shocks," to the outcome variable
  - no hurricanes
  - no volcanos

Conditions for Donor Pool Observations

- Untreated during observation period
    - no policy in the donor pool observations
- Unaffected by treatment
    - no spillover from treated observations
- Should have no large, “idiosyncratic shocks,” to the outcome variable
    - no hurricanes
    - no volcanos
- “Similar” to treated units to avoid interpolation bias
    - this is a little vague
    - hopefully the method does this for you

Conditions for Donor Pool Observations

- Untreated during observation period
  - no policy in the donor pool observations
- Unaffected by treatment
  - no spillover from treated observations
- Should have no large, "idiosyncratic shocks," to the outcome variable
  - no hurricanes
  - no volcanos
- "Similar" to treated units to avoid interpolation bias
  - this is a little vague
  - hopefully the method does this for you

Many of these are not unique to synthetic control

Outcomes and Set-up for Synthetic Control

Notation

- Outcome for treated unit if treated:
  Y_{it}^{I}
- Outcome for treated unit if untreated:
  Y_{it}^{N}

Outcomes and Set-up for Synthetic Control

Notation

- Outcome for treated unit if treated:
  Y_{it}^{I}
- Outcome for treated unit if untreated:
  Y_{it}^{N}
- We assume

$$Y_{it}^{I} = Y_{it}^{N} \text{ for any } t \leq T_0$$

Outcomes and Set-up for Synthetic Control

Notation

- Outcome for treated unit if treated: Y_{it}^{I}
- Outcome for treated unit if untreated: Y_{it}^{N}
- We assume

$$Y_{it}^{I} = Y_{it}^{N} \text{ for any } t \leq T_0$$

- $D_i = 1$ if i is ever treated, 0 otherwise
- Z_{it} are covariates

Outcomes and Set-up for Synthetic Control

Notation

- Outcome for treated unit if treated: Y_{it}^{I}
- Outcome for treated unit if untreated: Y_{it}^{N}
- We assume

$$Y_{it}^{I} = Y_{it}^{N} \text{ for any } t \leq T_0$$

- $D_i = 1$ if i is ever treated, 0 otherwise
- Z_{it} are covariates

Setting up the problem

- We want to know

Outcomes and Set-up for Synthetic Control

Notation

- Outcome for treated unit if treated: Y_{it}^{I}
- Outcome for treated unit if untreated: Y_{it}^{N}
- We assume

$$Y_{it}^{I} = Y_{it}^{N} \text{ for any } t \leq T_0$$

- $D_i = 1$ if i is ever treated, 0 otherwise
- Z_{it} are covariates

Setting up the problem

- We want to know difference between outcome in treated and untreated state:

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$$

Outcomes and Set-up for Synthetic Control

Notation

- Outcome for treated unit if treated: Y_{it}^{I}
- Outcome for treated unit if untreated: Y_{it}^{N}
- We assume

$$Y_{it}^{I} = Y_{it}^{N} \text{ for any } t \leq T_0$$

- $D_i = 1$ if i is ever treated, 0 otherwise
- Z_{it} are covariates

Setting up the problem

- We want to know difference between outcome in treated and untreated state:

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$$

- Note that the effect varies with time: α has a t subscript

Outcomes and Set-up for Synthetic Control

Notation

- Outcome for treated unit if treated: Y_{it}^{I}
- Outcome for treated unit if untreated: Y_{it}^{N}
- We assume

$$Y_{it}^{I} = Y_{it}^{N} \text{ for any } t \leq T_0$$

- $D_i = 1$ if i is ever treated, 0 otherwise
- Z_{it} are covariates

Setting up the problem

- We want to know difference between outcome in treated and untreated state:

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$$

- Note that the effect varies with time: α has a t subscript
- Does this differ from a diff-in-diff?

Outcomes and Set-up for Synthetic Control

Notation

- Outcome for treated unit if treated: Y_{it}^{I}
- Outcome for treated unit if untreated: Y_{it}^{N}
- We assume

$$Y_{it}^{I} = Y_{it}^{N} \text{ for any } t \leq T_0$$

- $D_i = 1$ if i is ever treated, 0 otherwise
- Z_{it} are covariates

Setting up the problem

- We want to know difference between outcome in treated and untreated state:

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$$

- Note that the effect varies with time: α has a t subscript
- Does this differ from a diff-in-diff?
- But! Problem: We only observe Y_{it}^{I}

Goal of Estimation

Goal in Estimation

- Goal: find

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$$

Goal in Estimation

- Goal: find

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$$

- We observe the treated unit in the treated state, so $Y_{it}^{I} = Y_{it}$
- Therefore,

$$\alpha_{it} = Y_{it} - Y_{it}^{N}$$

- What here is unknown?

Goal in Estimation

- Goal: find

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$$

- We observe the treated unit in the treated state, so $Y_{it}^{I} = Y_{it}$
- Therefore,

$$\alpha_{it} = Y_{it} - Y_{it}^{N}$$

- What here is unknown? Y_{it}^{N}

Goal in Estimation

- Goal: find

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$$

- We observe the treated unit in the treated state, so $Y_{it}^{I} = Y_{it}$
- Therefore,

$$\alpha_{it} = Y_{it} - Y_{it}^{N}$$

- What here is unknown? Y_{it}^{N}
- Fundamental question: how to estimate Y_{it}^{N}

Goal in Estimation

- Goal: find

$$\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$$

- We observe the treated unit in the treated state, so $Y_{it}^{I} = Y_{it}$
- Therefore,

$$\alpha_{it} = Y_{it} - Y_{it}^{N}$$

- What here is unknown? Y_{it}^{N}
- Fundamental question: how to estimate Y_{it}^{N}
- Intuition: approximate with a weighted average of untreated units
- In math

$$\hat{Y}_{it}^{N} = \sum_{j=2}^{J+1} w_j^{*} Y_{jt}$$

Goal in Estimation

- Goal: find

$$\alpha_{it} = Y^I_{it} - Y^N_{it}$$

- We observe the treated unit in the treated state, so $Y^I_{it} = Y_{it}$
- Therefore,

$$\alpha_{it} = Y_{it} - Y^N_{it}$$

- What here is unknown? Y^N_{it}
- Fundamental question: how to estimate Y^N_{it}
- Intuition: approximate with a weighted average of untreated units
- In math

$$\hat{Y}^N_{it} = \sum_{j=2}^{J+1} w^*_j Y_{jt}$$

Trick is to find w_j

Estimation

What's the Goal?

- In OLS, we minimize what to find the best fit line?

What's the Goal?

- In OLS, we minimize what to find the best fit line?
  - $\sum_{j=1}^{J} \epsilon_i^2$
  - In matrix language that $\epsilon'\epsilon$

What's the Goal?

- In OLS, we minimize what to find the best fit line?
  - $\sum_{j=1}^{J} \epsilon_i^2$
  - In matrix language that $\epsilon' \epsilon$
- In SC, we choose weights to minimize the difference between
  - the treated covariates and pre-treatment outcomes and
  - the donor pool's covariates and pre-treatment outcomes
- But remember that our optimal weights have **no** time dimension

Define What We Want To Match

We want X_0 to be like X_1

- X_1 is
  - pre-treatment covariates Z_1 and
  - pre-treatment outcomes $Y_{1,t<T_0}$ for the treated unit

$$X_1 \equiv \{Z_1, Y_{1,t<T_0}\}$$

Define What We Want To Match

We want X_0 to be like X_1

- X_1 is
  - pre-treatment covariates Z_1 and
  - pre-treatment outcomes $Y_{1,t<T_0}$ for the treated unit

$$X_1 \equiv \{Z_1, Y_{1,t<T_0}\}$$

- X_0 is
  - pre-treatment covariates Z_0 and
  - pre-treatment outcomes $Y_{0,t<T_0}$ for the untreated unit

$$X_0 \equiv \{Z_0, Y_{0,t<T_0}\}$$

Choosing Weights

- Choose weights $\{w_2, w_2, \ldots, w_{J+1}\} \in W$ to minimize

$$||X_1 - X_0 W||$$

Choosing Weights

- Choose weights $\{w_2, w_2, \ldots, w_{J+1}\} \in W$ to minimize

$$||X_1 - X_0 W||$$

- We have one weight for each of the J donor pool observations

What we are minimizing

Remember, Z are covariates, Y are pre-treatment outcomes

$$X_1 = \begin{pmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1r} \\ Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1T_0} \end{pmatrix},$$

What we are minimizing

Remember, Z are covariates, Y are pre-treatment outcomes

$$X_1 = \begin{pmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1r} \\ Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1T_0} \end{pmatrix}, X_0 = \begin{pmatrix} Z_{21} & \dots & Z_{J+1,1} \\ Z_{22} & \dots & Z_{J+1,2} \\ \vdots & \ddots & \vdots \\ Z_{2r} & \dots & Z_{J+1,r} \\ Y_{21} & \dots & Y_{J+1,1} \\ Y_{22} & \dots & Y_{J+1,2} \\ \vdots & \ddots & \vdots \\ Y_{2T_0} & \dots & Y_{J+1,T_0} \end{pmatrix},$$

What we are minimizing

Remember, Z are covariates, Y are pre-treatment outcomes

$$X_1 = \begin{pmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1r} \\ Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1T_0} \end{pmatrix}, X_0 = \begin{pmatrix} Z_{21} & \dots & Z_{J+1,1} \\ Z_{22} & \dots & Z_{J+1,2} \\ \vdots & \ddots & \vdots \\ Z_{2r} & \dots & Z_{J+1,r} \\ Y_{21} & \dots & Y_{J+1,1} \\ Y_{22} & \dots & Y_{J+1,2} \\ \vdots & \ddots & \vdots \\ Y_{2T_0} & \dots & Y_{J+1,T_0} \end{pmatrix}, W = \begin{pmatrix} w_2 \\ w_3 \\ \vdots \\ w_{J+1} \end{pmatrix}$$

Choosing Weights for Covariates

- We choose w_j so as to minimize $||X_1 - X_0 W||$

Choosing Weights for Covariates

- We choose w_j so as to minimize $||X_1 - X_0 W||$
- But
  - $||X_1 - X_0 W||$ is not one number
  - it is a vector
  - of as many numbers as you have Z and pre-treatment outcomes $Y_{t<T_0}$

Choosing Weights for Covariates

- We choose w_j so as to minimize $||X_1 - X_0 W||$
- But
  - $||X_1 - X_0 W||$ is not one number
  - it is a vector
  - of as many numbers as you have Z and pre-treatment outcomes $Y_{t<T_0}$
- $\rightarrow$ final choice: how to weight to give to each difference in pre-treatment covariate or outcome so you can minimize

Choosing the Weights for Covariates

- We find donor pool weights w_j **conditional** on a decision about how much importance to give to each element of X
- Specifically, each element of X gets a weight v_k
- A final frontier of this estimation

Example

Suppose that there is one covariate and two pre-treatment observations. The matrix looks like this

$$Z_1 = 5, Z_0 = \begin{pmatrix} 0 & 5 \end{pmatrix}, Y_{1,t<T_0} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, Y_{0,t<T_0} = \begin{pmatrix} 2 & 5 \\ 3 & 8 \end{pmatrix}$$

Example

Suppose that there is one covariate and two pre-treatment observations. The matrix looks like this

$$Z_1 = 5, Z_0 = \begin{pmatrix} 0 & 5 \end{pmatrix}, Y_{1,t<T_0} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, Y_{0,t<T_0} = \begin{pmatrix} 2 & 5 \\ 3 & 8 \end{pmatrix}$$

$$X_1 = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix}, X_0 = \begin{pmatrix} 0 & 5 \\ 2 & 5 \\ 3 & 8 \end{pmatrix},$$

Example

Suppose that there is one covariate and two pre-treatment observations. The matrix looks like this

$$Z_1 = 5, Z_0 = \begin{pmatrix} 0 & 5 \end{pmatrix}, Y_{1,t<T_0} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, Y_{0,t<T_0} = \begin{pmatrix} 2 & 5 \\ 3 & 8 \end{pmatrix}$$

$$X_1 = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix}, X_0 = \begin{pmatrix} 0 & 5 \\ 2 & 5 \\ 3 & 8 \end{pmatrix}, W = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

Example

Suppose that there is one covariate and two pre-treatment observations. The matrix looks like this

$$Z_1 = 5, Z_0 = \begin{pmatrix} 0 & 5 \end{pmatrix}, Y_{1,t<T_0} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, Y_{0,t<T_0} = \begin{pmatrix} 2 & 5 \\ 3 & 8 \end{pmatrix}$$

$$X_1 = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix}, X_0 = \begin{pmatrix} 0 & 5 \\ 2 & 5 \\ 3 & 8 \end{pmatrix}, W = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

$$||X_1 - X_0 W|| = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 0 & 5 \\ 2 & 5 \\ 3 & 8 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 0w_1 + 5w_2 \\ 2w_1 + 5w_2 \\ 3w_1 + 8w_2 \end{pmatrix} = \begin{pmatrix} 5 - (0w_1 + 5w_2) \\ 2 - (2w_1 + 5w_2) \\ 3 - (3w_1 + 8w_2) \end{pmatrix}$$

And the Last Bit

- $||X_1 - X_0 W||$ is a vector, which means a list of numbers
- How do you minimize a vector?

And the Last Bit

- $||X_1 - X_0W||$ is a vector, which means a list of numbers
- How do you minimize a vector? You don't
- We do know how to choose donor weights to minimize one number

And the Last Bit

- $||X_1 - X_0W||$ is a vector, which means a list of numbers
- How do you minimize a vector? You don't
- We do know how to choose donor weights to minimize one number
- $\rightarrow$ make the vector one number by adding up its parts

And the Last Bit

- $||X_1 - X_0W||$ is a vector, which means a list of numbers
- How do you minimize a vector? You don't
- We do know how to choose donor weights to minimize one number
- $\rightarrow$ make the vector one number by adding up its parts
- Have to decide how much we care about different parts of the diversion from the treated outcome to add them up
- The sum of all the parts of the vector is the mean squared error of the estimate
    - $\text{MSE} = ||X_1 - X_0W||v$, where v is yet another weighting matrix

How do you choose v?

A variety of options

- Give equal weight to all elements of X
- So that the pre-intervention difference in Y is minimized
- To minimize error in the final estimation (Abadie et al do this in another paper)
- Cross-validation in Germany paper:
  - find W for the first half of the pre-treatment era
  - choose v such that $||X_1 - X_0W||v$ is minimized in the second half of the pre-treatment period
  - if there are multiple possible W, you can see which one gives the lowest MSPE in the second pre-treatment period

How do you choose v?

A variety of options

- Give equal weight to all elements of X
- So that the pre-intervention difference in Y is minimized
- To minimize error in the final estimation (Abadie et al do this in another paper)
- Cross-validation in Germany paper:
  - find W for the first half of the pre-treatment era
  - choose v such that $||X_1 - X_0W||v$ is minimized in the second half of the pre-treatment period
  - if there are multiple possible W, you can see which one gives the lowest MSPE in the second pre-treatment period

Note that $||X_1 - X_0W||v$ is the Mean Squared Prediction Error: MSPE

Re-capping Assumptions

- No effect of treatment on the untreated
- The treated unit would have had the untreated outcome in the absence of treatment
- Treated observation is in the convex hull of the donor pool

The Necessity of a Convex Hull

What is a Convex Hull?

- Key requirement
  - X_1 is in the convex hull of X_0
  - In words: the treated outcome's matching variables are in the convex hull of all donor pool Y and pre-treatment Z

What is a Convex Hull?

- Key requirement
  - X_1 is in the convex hull of X_0
  - In words: the treated outcome's matching variables are in the convex hull of all donor pool Y and pre-treatment Z
- What is that?
  - In general, the convex hull of A is the "smallest convex set that contains A"

What is a Convex Hull?

- Key requirement
  - X_1 is in the convex hull of X_0
  - In words: the treated outcome's matching variables are in the convex hull of all donor pool Y and pre-treatment Z
- What is that?
  - In general, the convex hull of A is the "smallest convex set that contains A"
  - Think of a set of three points (x, y)
  - What is the shape of the convex hull of three points?

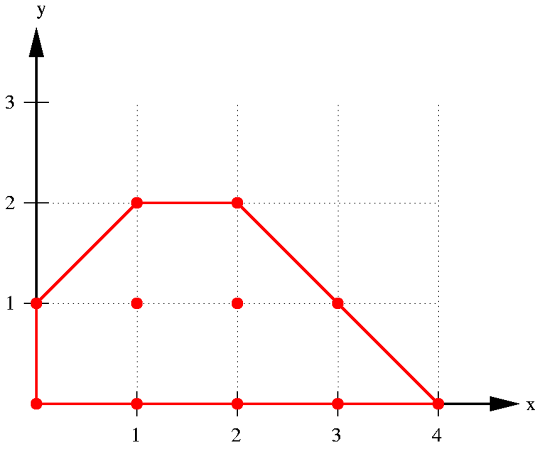
What is a Convex Hull?

- Key requirement
  - X_1 is in the convex hull of X_0
  - In words: the treated outcome's matching variables are in the convex hull of all donor pool Y and pre-treatment Z
- What is that?
  - In general, the convex hull of A is the "smallest convex set that contains A"
  - Think of a set of three points (x, y)
  - What is the shape of the convex hull of three points? triangle

What is a Convex Hull?

- Key requirement
  - X_1 is in the convex hull of X_0
  - In words: the treated outcome's matching variables are in the convex hull of all donor pool Y and pre-treatment Z
- What is that?
  - In general, the convex hull of A is the "smallest convex set that contains A"
  - Think of a set of three points (x, y)
  - What is the shape of the convex hull of three points? triangle
  - If you are inside the triangle, you are in the convex hull
  - If you are outside the triangle, you are outside the convex hull

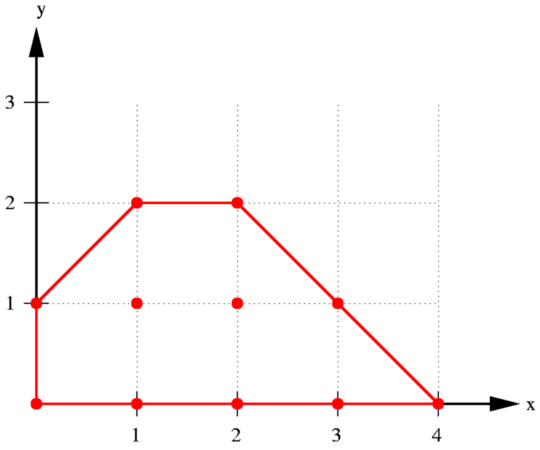
Convex Hull Example



- Red dots are the set of points $x \in X$
- The red line is the outline of the convex hull
- Is $(2, 1)$ inside the convex hull?

Note: source is http://en.wikipedia.org/wiki/File:Convex_hull.png

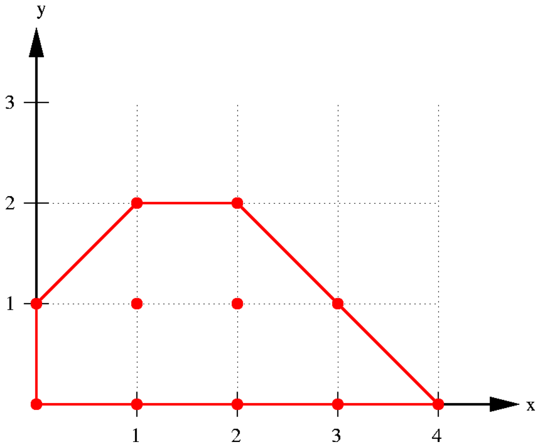
Convex Hull Example



- Red dots are the set of points $x \in X$
- The red line is the outline of the convex hull
- Is $(2, 1)$ inside the convex hull? yes

Note: source is http://en.wikipedia.org/wiki/File:Convex_hull.png

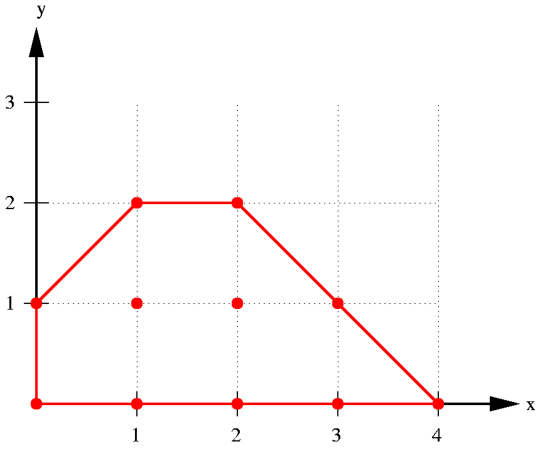
Convex Hull Example



- Red dots are the set of points $x \in X$
- The red line is the outline of the convex hull
- Is $(2, 1)$ inside the convex hull? yes
- If $(2, 1)$ is the treated point you can use a synthetic control

Note: source is http://en.wikipedia.org/wiki/File:Convex_hull.png

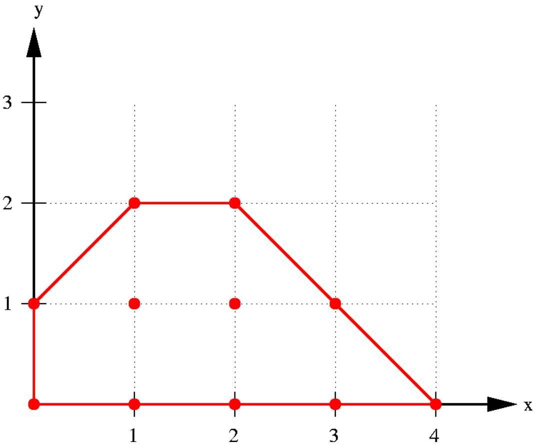
Convex Hull Example



- Red dots are the set of points $x \in X$
- The red line is the outline of the convex hull
- Is $(2, 1)$ inside the convex hull? yes
- If $(2, 1)$ is the treated point you can use a synthetic control
- Is $(3, 3)$ inside the convex hull?

Note: source is http://en.wikipedia.org/wiki/File:Convex_hull.png

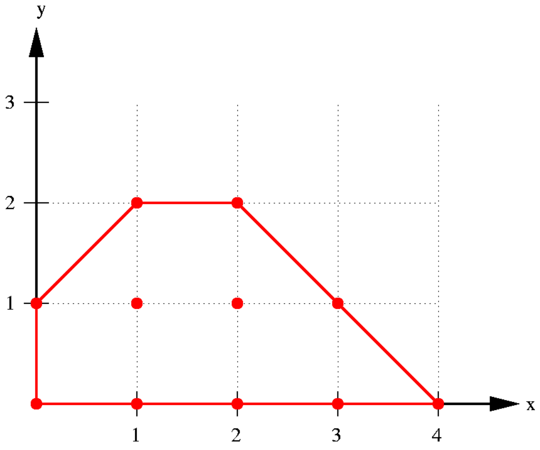
Convex Hull Example



- Red dots are the set of points $x \in X$
- The red line is the outline of the convex hull
- Is $(2, 1)$ inside the convex hull? yes
- If $(2, 1)$ is the treated point you can use a synthetic control
- Is $(3, 3)$ inside the convex hull? no

Note: source is http://en.wikipedia.org/wiki/File:Convex_hull.png

Convex Hull Example



- Red dots are the set of points $x \in X$
- The red line is the outline of the convex hull
- Is $(2,1)$ inside the convex hull? yes
- If $(2,1)$ is the treated point you can use a synthetic control
- Is $(3,3)$ inside the convex hull? no
- If $(3,3)$ is the treated point you **cannot** use a synthetic control

Note: source is http://en.wikipedia.org/wiki/File:Convex_hull.png

What Has a Convex Hull?

- Think of a treatment and observations where the donor pool would not form a convex hull.

What Has a Convex Hull?

- Think of a treatment and observations where the donor pool would not form a convex hull.
  - Impact of an additional billion dollars on Bill Gates's charitable giving

What Has a Convex Hull?

- Think of a treatment and observations where the donor pool would not form a convex hull.
  - Impact of an additional billion dollars on Bill Gates's charitable giving
  - Impact of additional billionaires on building height in New York

What Has a Convex Hull?

- Think of a treatment and observations where the donor pool would not form a convex hull.
  - Impact of an additional billion dollars on Bill Gates's charitable giving
  - Impact of additional billionaires on building height in New York
- Sufficient condition for having donor pool observations in the convex hull
  - the "number of pre-intervention periods is large relative to the scale of the transitory shocks"

What Has a Convex Hull?

- Think of a treatment and observations where the donor pool would not form a convex hull.
  - Impact of an additional billion dollars on Bill Gates's charitable giving
  - Impact of additional billionaires on building height in New York
- Sufficient condition for having donor pool observations in the convex hull
  - the "number of pre-intervention periods is large relative to the scale of the transitory shocks"
- Assuming that the donor pool lies in the convex hull is equivalent to assuming

$$X_1 - \sum_{j=2}^{J+1} w_j X_0 \equiv 0$$

  in words: pre-treatment outcome for treated observation can be re-created via a weighted average of untreated observations
- The convex hull assumption is sort of testable

Statistical Inference

Deciding Whether the Result is More Than Random Chance

- We always begin statistics with “descriptive statistics”
  - examples?

Deciding Whether the Result is More Than Random Chance

- We always begin statistics with “descriptive statistics”
    - examples? mean, median, standard deviation
    - these statistics are “properties of the observed data”

Deciding Whether the Result is More Than Random Chance

- We always begin statistics with "descriptive statistics"
  - examples? mean, median, standard deviation
  - these statistics are "properties of the observed data"
- But sometimes we want to know how these observed data relate to the larger population from which they are sampled
- We turn in this case to "inferential statistics"

Deciding Whether the Result is More Than Random Chance

- We always begin statistics with "descriptive statistics"
  - examples? mean, median, standard deviation
  - these statistics are "properties of the observed data"
- But sometimes we want to know how these observed data relate to the larger population from which they are sampled
- We turn in this case to "inferential statistics"
  - regression coefficient, which approximates parameter of interest
  - confidence interval

Deciding Whether the Result is More Than Random Chance

- We always begin statistics with “descriptive statistics”
  - examples? mean, median, standard deviation
  - these statistics are “properties of the observed data”
- But sometimes we want to know how these observed data relate to the larger population from which they are sampled
- We turn in this case to “inferential statistics”
  - regression coefficient, which approximates parameter of interest
  - confidence interval
- How do we make statistical inference in the case of synthetic control?

Thank you “statistical inference” in Wikipedia.

Statistical Interference in Synthetic Control

- A very open question
- Check back in 5 to 10 years

Statistical Interference in Synthetic Control

- A very open question
- Check back in 5 to 10 years
- But we have some interim things we can do

One Way to Evaluate Error in Synthetic Control

RMSPE: Root Mean Squared Prediction Error

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

This equation in words

- for each period before the treatment, $t < T_0$
- find squared difference between treated outcome, Y_{1t} , and synthetic control, $\sum_{j=2}^{J+1} w_j^* Y_{jt}$
- add up for all pre-treatment years and take square root

Evaluating Root Mean Squared Prediction Error

- Is a bigger or smaller RMSPE a good sign?
- First, think pre-treatment

Evaluating Root Mean Squared Prediction Error

- Is a bigger or smaller RMSPE a good sign?
- First, think pre-treatment
  - smaller is better
  - a synthetic control a better match
- Then post-treatment

Evaluating Root Mean Squared Prediction Error

- Is a bigger or smaller RMSPE a good sign?
- First, think pre-treatment
  - smaller is better
  - a synthetic control a better match
- Then post-treatment
  - bigger RMSPE is evidence of larger impact of treatment
  - more deviation from synthetic control

Evaluating Root Mean Squared Prediction Error

- Is a bigger or smaller RMSPE a good sign?
- First, think pre-treatment
  - smaller is better
  - a synthetic control a better match
- Then post-treatment
  - bigger RMSPE is evidence of larger impact of treatment
  - more deviation from synthetic control
- Then a big value for post / pre treatment RMSPE may be useful

Two More Inferential Methods

1. In time placebo

Two More Inferential Methods

1. In time placebo
   - Pretend the evaluation took place earlier, or some $t < T_0$
   - Do synthetic control estimation
   - Do we find results after the fake treatment?

Two More Inferential Methods

1. In time placebo
   - Pretend the evaluation took place earlier, or some $t < T_0$
   - Do synthetic control estimation
   - Do we find results after the fake treatment?
2. In space placebo
   - Pretend that another observation is the treated one

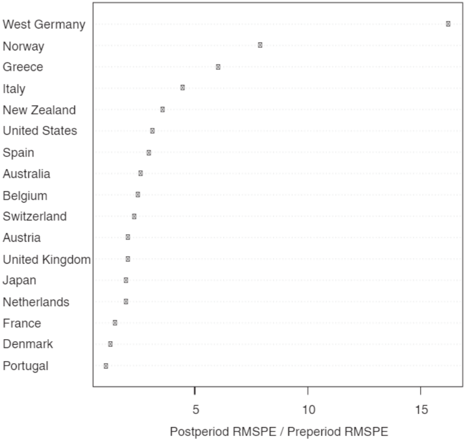
Two More Inferential Methods

1. In time placebo
   - Pretend the evaluation took place earlier, or some $t < T_0$
   - Do synthetic control estimation
   - Do we find results after the fake treatment?

2. In space placebo
   - Pretend that another observation is the treated one
   - Do the synthetic control
   - Is post-treatment RMSPE for treated obs bigger than for other obs?

In Space Placebo

FIGURE 5 Ratio of Postreunification RMSPE to Prereunification RMSPE: West Germany and Control Countries



- Do synthetic control method for each country as if it were treated
- Find pre- and post-treatment RMSPE
- This is goodness of fit before – what you want
- And badness of fit afterward – what you see if there is an effect
- Chart reports this ratio
- West Germany is an outlier

California Tobacco Tax Example

Tobacco tax in California

- In 1998 CA passed Prop 99: tobacco tax increase + funding for tobacco control programs
- Did Prop. 99 decrease smoking?
- What endogeneity issues could be concerning?

Tobacco tax in California

- In 1998 CA passed Prop 99: tobacco tax increase + funding for tobacco control programs
- Did Prop. 99 decrease smoking?
- What endogeneity issues could be concerning?
- Use state level data on smoking and state characteristics
- Who is in the synthetic control?

Tobacco tax in California

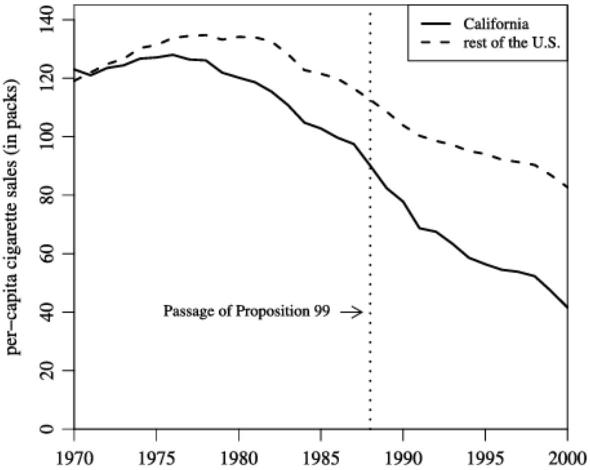
- In 1998 CA passed Prop 99: tobacco tax increase + funding for tobacco control programs
- Did Prop. 99 decrease smoking?
- What endogeneity issues could be concerning?
- Use state level data on smoking and state characteristics
- Who is in the synthetic control?
- Utah (0.33), Nevada (0.23), Montana (0.22), Colorado (0.16), CT (0.07)

Abadie, Diamond and Hainmueller, 2010

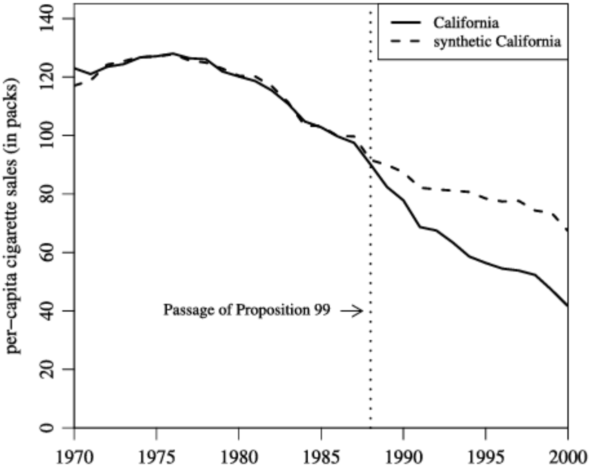
Things You Should Expect in a Synthetic Control Paper

1. Comparison without synthetic control
2. Comparison with synthetic control
3. Covariates with and without synthetic control
4. And sometimes even covariates on which we don't match
5. Some kind of inference

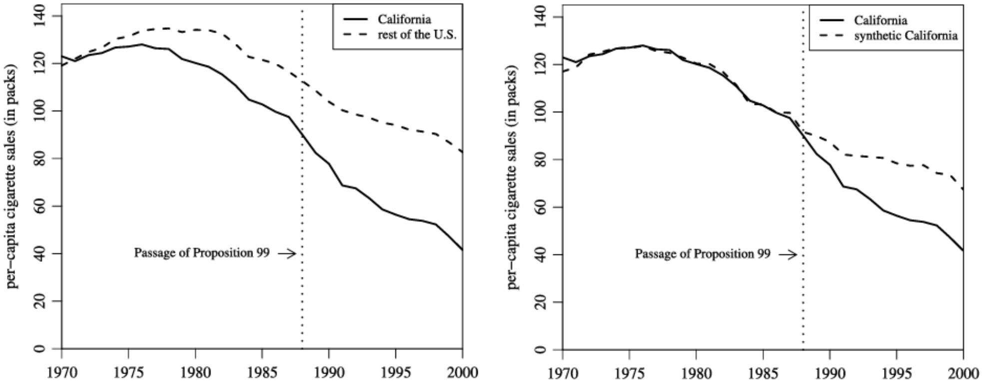
Prop. 99: Comparison without Synthetic Control



Prop. 99: Comparison With Synthetic Control



Prop. 99: Both Comparisons



Prop. 99: Covariates With Synthetic Control

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

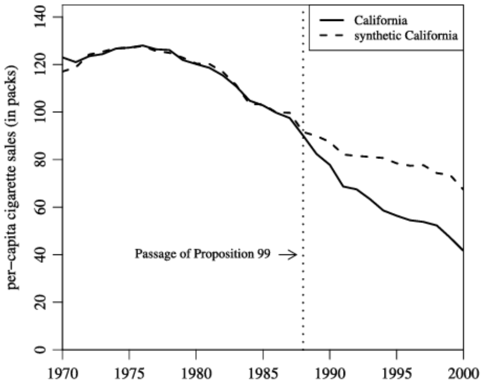
Prop. 99: Covariates With Synthetic Control

Table 1. Cigarette sales predictor means

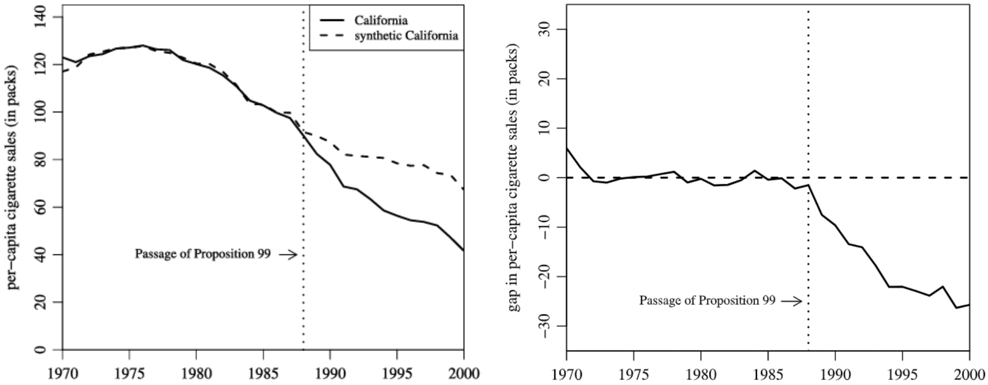
Variables	California Real	California Synthetic	Average of 38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15–24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

These are targeted covariates – we can also compare untargeted covariates

Prop. 99: Starting Inference

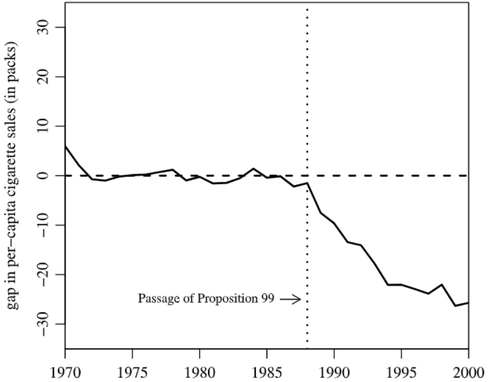


Prop. 99: Starting Inference



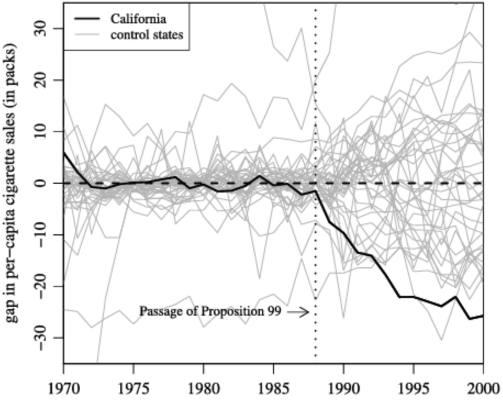
Prop. 99: Inference

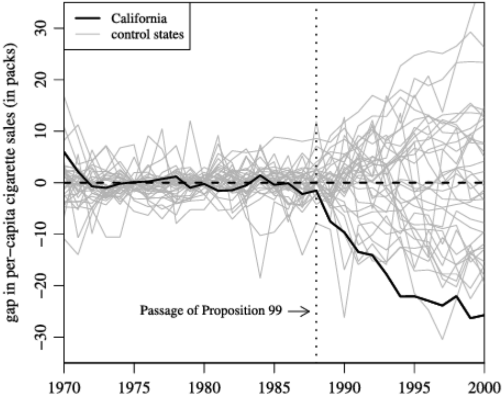
- Repeat analysis for California
- But with each donor pool state as "treated"
- One gray line for each placebo state
- If Prop. 99 has a real effect, what should the other gray lines look like?



Left panel has all states; right panel limits to 34 states with pre-treatment RMPSE $<$ 20 times CA's

Prop. 99: Possibly Stronger Inference





Left panel has all states; right panel limits to 34 states with pre-treatment RMPSE $<$ 20 times CA's

Prop. 99: Alternative Inference

- Divide squared treatment-control difference after treatment by
- squared treatment-control difference before treatment
- If Prop. 99 has a real effect, should this be big or small?

Prop. 99: Alternative Inference

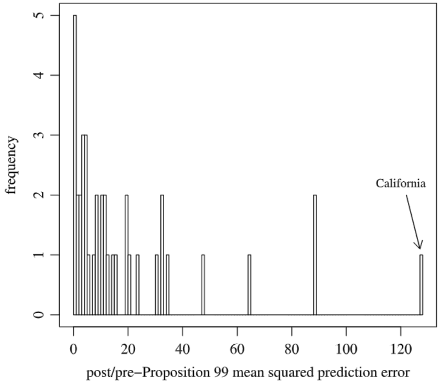
- Divide squared treatment-control difference after treatment by
- squared treatment-control difference before treatment
- If Prop. 99 has a real effect, should this be big or small?

$$\text{MSPE}_{j=s} = \frac{\frac{1}{T_0}\sum_{T=0}^{T_0}(Y_{j=s,t} - Y_{j\neq s,t})^2}{\frac{1}{(T-(T_0+1))}\sum_{T=T_0+1}^{T}(Y_{j=s,t} - Y_{j\neq s,t})^2}$$

Prop. 99: Alternative Inference

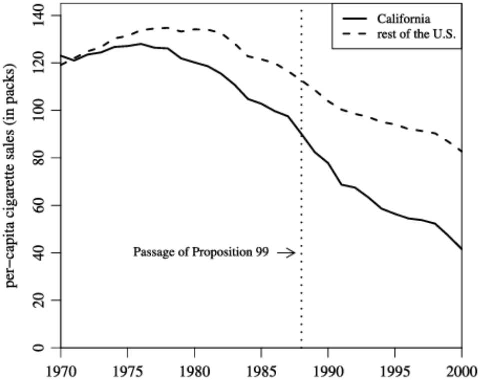
- Divide squared treatment-control difference after treatment by
- squared treatment-control difference before treatment
- If Prop. 99 has a real effect, should this be big or small?

$$\text{MSPE}_{j=s} = \frac{\frac{1}{T_0}\sum_{T=0}^{T_0}(Y_{j=s,t} - Y_{j\neq s,t})^2}{\frac{1}{(T-(T_0+1))}\sum_{T=T_0+1}^{T}(Y_{j=s,t} - Y_{j\neq s,t})^2}$$



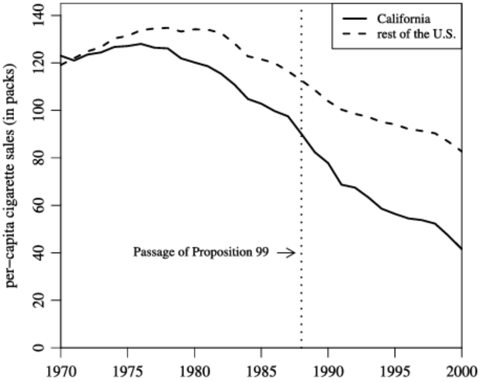
Prop. 99: Comparison

No synthetic control

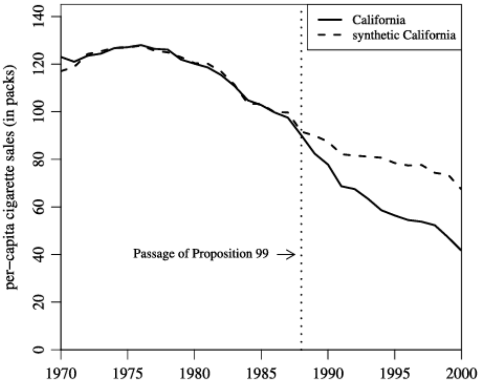


Prop. 99: Comparison

No synthetic control



With synthetic control



Upcoming: The End is in Sight

- Lecture 12: In-class workshop and how to explain a causal strategy
- Lecture 13: Presentations
- Lecture 14: Presentations