Evaluating Policy and Quantifying Uncertainty with Few (or One) Treated Unit(s): An Introduction to Synthetic Control Methods and Falsification Analyses

Sydney Kahmann

Department of Statistics
University of California, Los Angeles

skahmann@ucla.edu

Based on joint work with Erin Hartman, Jorja Leap, and P. Jeffrey Brantingham.
Link to paper

February 19, 2021

Outline

- An introduction to causal inference
- Policy evaluation as a causal problem
  - Observational data
  - Few treated units
  - Selection bias
  - Time-varying trend
  - Limited controls
- Methodology:
  - Synthetic control methods
  - Falsification analyses
- Application:
  - Community Safety Partnership

An Overview of Causal Inference for Policy Evaluation

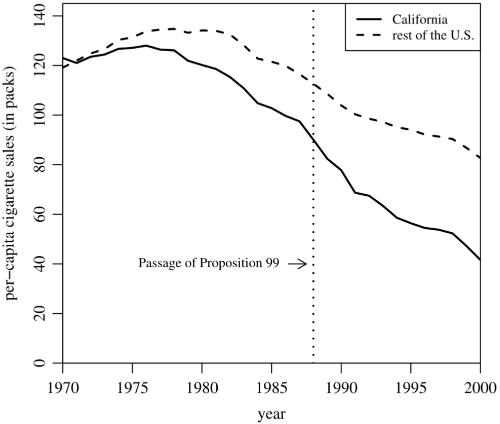

Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Abadie, Diamond, and Hainmueller (2010).

- Did the law cause sales to decrease more than we would have expected otherwise?
- Treatment effect = (observed post-treatment sales in CA) - (what sales would have been in CA during the post-treatment period had treatment not been implemented)

An Overview of Causal Inference for Policy Evaluation

- Did the law cause sales to decrease more than we would have expected otherwise?
- Treatment effect = (observed post-treatment sales in CA) - (what sales would have been in CA during the post-treatment period had treatment not been implemented)

For a given unit at a given time, we can only observe one possible state.

- Therefore, each unit has a set of **Potential Outcomes**
- The observed state: **Factual**
- The unobserved state: **Counterfactual**

An Overview of Causal Inference for Policy Evaluation

- Units indexed by $i \in (1, \ldots, N)$
- Time periods indexed by $t \in (1, \ldots, T)$
- Treatment implemented at time $t = T_0$
- Define Y as the outcome of interest *(i.e. cigarette sales)*
- Define D as the treatment assignment mechanism, D $=$ 1 indicates units receiving treatment. *(i.e. anti-smoking law)*

An Overview of Causal Inference for Policy Evaluation

- Units indexed by $i \in (1, \ldots, N)$
- Time periods indexed by $t \in (1, \ldots, T)$
- Treatment implemented at time $t = T_0$
- Define Y as the outcome of interest *(i.e. cigarette sales)*
- Define D as the treatment assignment mechanism, D = 1 indicates units receiving treatment. *(i.e. anti-smoking law)*
- An individual's set of potential outcomes is

$$\{Y_{it}(D = 0), Y_{it}(D = 1)\}$$

An Overview of Causal Inference for Policy Evaluation

- Units indexed by $i \in (1, \ldots, N)$
- Time periods indexed by $t \in (1, \ldots, T)$
- Treatment implemented at time $t = T_0$
- Define Y as the outcome of interest *(i.e. cigarette sales)*
- Define D as the treatment assignment mechanism, D = 1 indicates units receiving treatment. *(i.e. anti-smoking law)*
- An individual's set of potential outcomes is

$$\{Y_{it}(D = 0), Y_{it}(D = 1)\}$$

- Treatment effect:

$$\tau_{it} = Y_{it}(1) - Y_{it}(0)$$

Estimand: Average Treatment Effect on the Treated (ATT)

$$ATT = \mathbb{E}[Y_{it}(1) - Y_{it}(0) | D_i = 1]$$

An Overview of Causal Inference for Policy Evaluation

Estimand: Average Treatment Effect on the Treated (ATT)

$$ATT = \mathbb{E}[Y_{it}(1) - Y_{it}(0) | D_i = 1]$$

ATT: average difference in *observed cigarette sales* in CA and *what cigarette sales would have been* in CA had Prop 99 not passed

An Overview of Causal Inference for Policy Evaluation

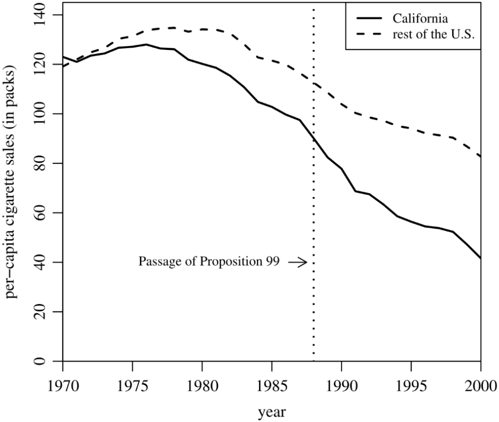

Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Abadie, Diamond, and Hainmueller (2010).

An Overview of Causal Inference for Policy Evaluation

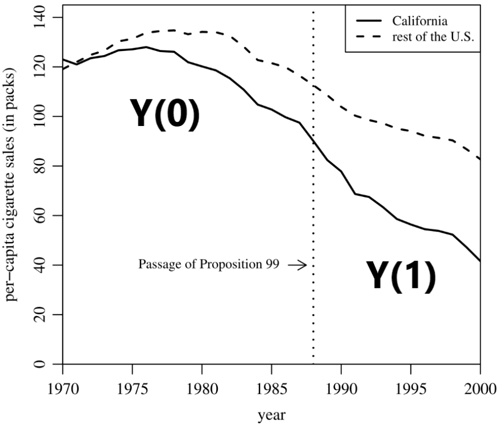

Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Abadie, Diamond, and Hainmueller (2010).

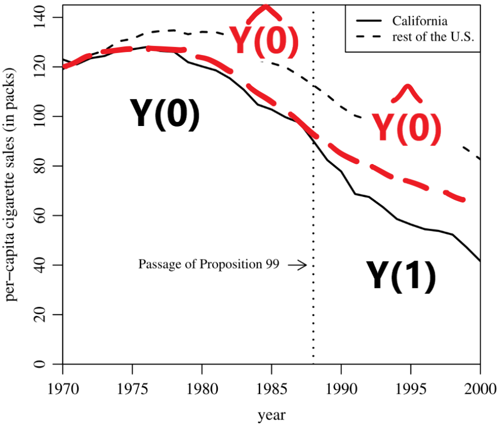

Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Abadie, Diamond, and Hainmueller (2010).

Synthetic Control Method

Goal: use set of weights, w_i^*, on the control units to construct a combination of units which is a good estimate of baseline CA, $Y_{it}(0)$

Synthetic Control Method

Goal: use set of weights, w_i^*, on the control units to construct a combination of units which is a good estimate of baseline CA, $Y_{it}(0)$

$$\widehat{ATT}_t = \underbrace{\frac{1}{n_t}\sum_{i \le n_t} Y_{it}(1)}_{\text{Observed Treated Behavior}} - \underbrace{\sum_{i > n_t} w_i^* Y_{it}(0)}_{\text{SCM Estimated Control Behavior}}$$

$\widehat{ATT}$ estimate: average post-treatment difference

Synthetic Control Method

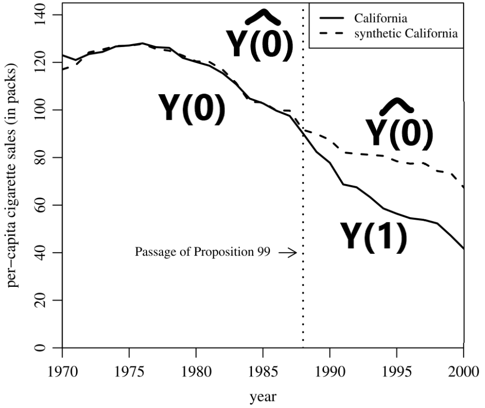

Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

- Data-driven combination of controls
- Accounts for temporal confounding
- Augmented SCM adjusts for remaining pre-treatment bias

Abadie, Diamond, and Hainmueller (2010).

Falsification Testing

- Goal: evaluate the credibility of the SCM model and final results
- True effect unknown and assumptions are untestable: require both potential outcomes
- Can check observable implications of the assumption

Falsification Testing

- Goal: evaluate the credibility of the SCM model and final results
- True effect unknown and assumptions are untestable: require both potential outcomes
- Can check observable implications of the assumption

Example: "As-if" randomization of treatment

Falsification Testing

- Goal: evaluate the credibility of the SCM model and final results
- True effect unknown and assumptions are untestable: require both potential outcomes
- Can check observable implications of the assumption

Example: "As-if" randomization of treatment

- Conditional ignorability assumption: $\{Y(0), Y(1)\} \perp\!\!\!\perp D | X$
- Observable Implication:

Falsification Testing

- Goal: evaluate the credibility of the SCM model and final results
- True effect unknown and assumptions are untestable: require both potential outcomes
- Can check observable implications of the assumption

Example: "As-if" randomization of treatment

- Conditional ignorability assumption: $\{Y(0), Y(1)\} \perp\!\!\!\perp D | X$
- Observable Implication:
  - "As-if" randomization in covariates across treatment groups
  - Mean balance between treatment and control groups in covariates

Falsification Testing

Simplest SCM falsification test:

- Pre-treatment balance between $Y(0)$ and $\widehat{Y(0)}$

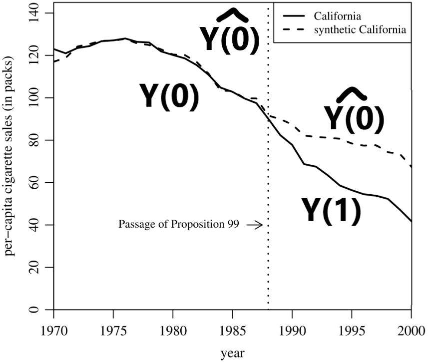

Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Abadie, Diamond, and Hainmueller (2010).

Falsification Testing

Goals:

- Assess model fit without being influenced by final results
- Evaluate the contextual significance of the results

"Placebo Effect" - common metric for evaluation:

- Placebo test: estimate effect where none should exist
    - Pre-treatment period
    - Control units
    - Placebo outcomes

Falsification Testing in Practice

Model Specification:

- Does the set of variables reflect observed control behavior?

Falsification Testing in Practice

Model Specification:

- Does the set of variables reflect observed control behavior?

Temporal Confounding:

- Are there potential anticipation effects or pre-treatment shocks?

Falsification Testing in Practice

Model Specification:

- Does the set of variables reflect observed control behavior?

Temporal Confounding:

- Are there potential anticipation effects or pre-treatment shocks?

Contextual Significance:

- Are the results significant compared to "effects" in control units?

Falsification Testing in Practice

Model Specification:

- Does the set of variables reflect observed control behavior?

Temporal Confounding:

- Are there potential anticipation effects or pre-treatment shocks?

Contextual Significance:

- Are the results significant compared to "effects" in control units?

Additional Approaches in Paper:

- Spillover effects: Are the controls receiving some form of treatment?
- Robustness: Are the results robust to alternative model specifications?

Community Safety Partnership

- Launched in late 2011 in two South LA public housing developments
- Shift from paramilitary to community policing
- Specially-trained CSP officers support and develop community and youth programs to improve quality of life and reduce violent crime

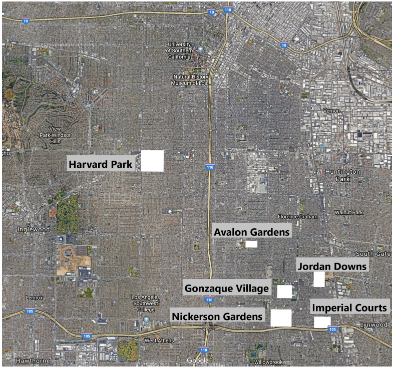

Figure: A Google Earth view of the region of interest with CSP regions labelled.

Community Safety Partnership

- Units: Census Block Group (space), Semester (time)
- Treatment Date: 2012, Period of Study: 2007-2017
- Three outcomes:
  - Violent crime*
  - Burglary
  - Quality of life

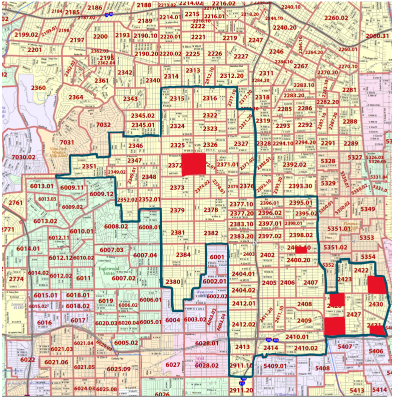

Figure: South Bureau in terms of Census Tract boundaries. Region of study outlined in blue. CSP public housing developments in red.

Model Specification Test

Does this model specification reflect observed control behavior?

- Split pre-treatment data into train, test sets using $2/3 : 1/3$ rule
- Psuedo assign treatment: 2010.5
- Goal: estimate negligible placebo effects without "p-hacking"

Model Specification Test

Does this model specification reflect observed control behavior?

- Split pre-treatment data into train, test sets using 2/3 : 1/3 rule
- Psuedo assign treatment: 2010.5
- Goal: estimate negligible placebo effects without “p-hacking”

Outcome	Pre-T. Average	Placebo ATT	Bias
Violent Crime	31.28	-2.55	8 %
Burglary	15.56	5.11	33 %
Quality of Life	194.94	-6.02	3 %

Table: Estimated placebo impact of CSP.

Temporal Confounding Test

Are the results highly influenced by a few data points?
Are there potential anticipation effects or pre-treatment shocks?

- Psuedo-assign treatment before actual implementation
- Goal: psuedo-implementation models follow the estimated model

Temporal Confounding Test

Are the results highly influenced by a few data points?
Are there potential anticipation effects or pre-treatment shocks?

- Psuedo-assign treatment before actual implementation
- Goal: psuedo-implementation models follow the estimated model

In the absence of confounding events during the pre-treatment period, we would expect to see the pseudo-implementation ATT_t estimates closely follow the ATT_t estimated from the true treatment period.

Temporal Confounding Test

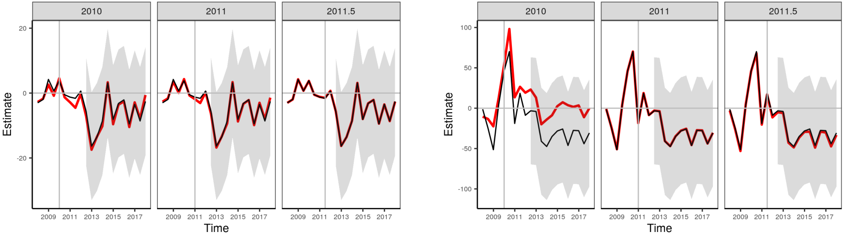

Figure: Left: violent crime. Right: quality of life.
Psuedo ATT_t estimates (red), estimated ATT_t (black with shaded bounds).
Psuedo-implementation date in panel title.

Contextual Significance Test

Are the results significant compared to "effects" in control units?

- Construct a distribution of placebo effects using the control units
- For each control, "assign" treatment to the control and shift the original treated unit to the donor pool
- Estimate the effect of CSP on the psuedo-treated control unit
- P-value: proportion of control models with higher RMSPEs than the treated model after removing poorly fitted control models

Contextual Significance Test

Are the results significant compared to “effects” in control units?

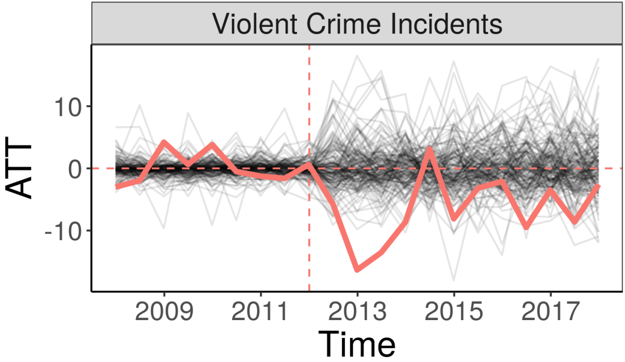

Figure: The comparative scale of the ATT_t effect versus the distribution of placebo effects. The p-value of 0.39 ($\frac{61}{157}$) is insignificant.

Falsification Results for Violent Crime

Model Specification:

- Chosen variables are a good estimate of pre-treatment behavior
- 8 % Bias (placebo effect / pre-treatment average)

Pre-Treatment Confounding:

- Psuedo implementation models do not estimate substantively different effects from the final model

Falsification Results for Violent Crime

Model Specification:

- Chosen variables are a good estimate of pre-treatment behavior
- 8 % Bias (placebo effect / pre-treatment average)

Pre-Treatment Confounding:

- Psuedo implementation models do not estimate substantively different effects from the final model

Contextual Significance:

- Estimated effect is not significantly different from distribution of placebo effects

CSP Results: Violent Crime

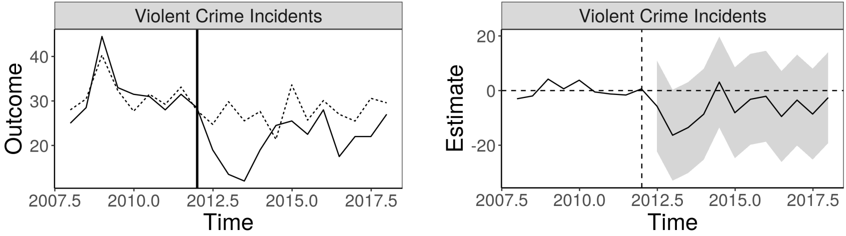

Figure: Left: violent crime trajectories for the treated (solid) vs. SCM (dashed). Right: the ATT_t estimates with shaded conformal inference bounds.

CSP Results: Violent Crime

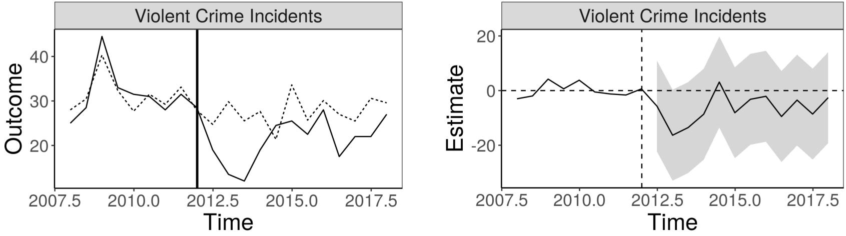

Figure: Left: violent crime trajectories for the treated (solid) vs. SCM (dashed). Right: the ATT_t estimates with shaded conformal inference bounds.

- ATT: -6.55 violent crimes per unit per six month period
- Average reduction of 21% compared to pre-treatment levels

Causal Inference for Policy Evaluation

- Why is policy evaluation a difficult causal problem?
  - Observational data
  - Few treated units
  - Selection bias
  - Time-varying trend
  - Limited controls
- Methodology:
  - Synthetic control methods
  - Falsification testing
  - Placebo tests
- Application:
  - Community Safety Partnership

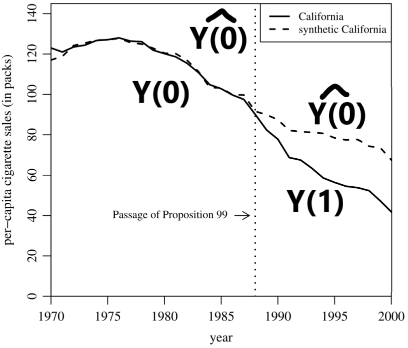

Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Abadie, Diamond, and Hainmueller (2010).

For additional examples of falsification tests: **link to paper**