Class 12: Synthetic Control Designs

POLS 6388: Causal Inference

Michael Kistner

Department of Political Science, University of Houston

April 08, 2025

Goals of Today's Class

1. Motivate + understand the logic of the synthetic control design
2. Learn how to implement the traditional SC design ($n^{Treat} = 1$)
3. Understand how to quantify uncertainty using randomization inference
4. Discuss extensions with multiple treated units (generalized synthetic control)

Introducing the Synthetic Control Design

A Different Panel Data Approach

When we have panel data + *many* treated AND *many* non-treated units, DiD is a powerful tool.

- Depends critically on the parallel trend assumption (PTA)
- What if the PTA doesn't hold for our research problem?

Another complication: How do we evaluate the effects of infrequent/rare events?

- Small n^{Treat}, perhaps only one
- We could try an interrupted time series design. But modeling the counterfactual without any control units is tough!
- If we have non-treated units, we should use them. What's the best way?

An Example: The Indiana Opioid RX Restriction

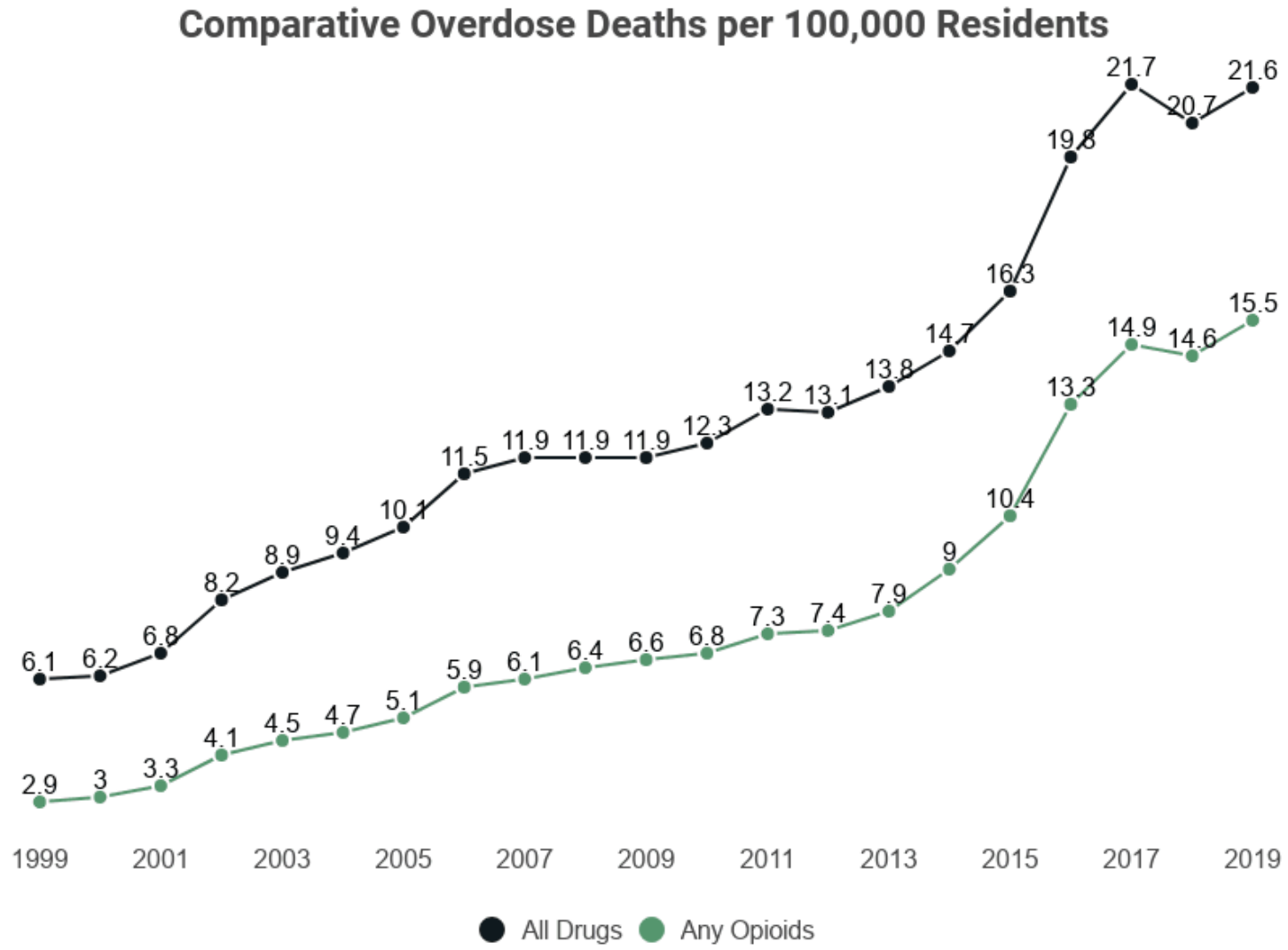
In 2012, Indiana implemented a policy that limited opioid prescriptions to 7 days for first time patients.

- Suppose Indiana is the only state that passed such a policy, but we want to evaluate the impact of the policy on opioid deaths.

How can we evaluate the counterfactual of Indiana without a RX restriction?

- Ideally, we'd find similar states to Indiana that *didn't* pass a prescription restriction, and compare opioid death trends.
- **A concern:** different states were affected by the opioid epidemic differentially (violating the PTA)

An Escalating Crisis from 2000 Onwards...



…with Disproportionate Impact on Some States

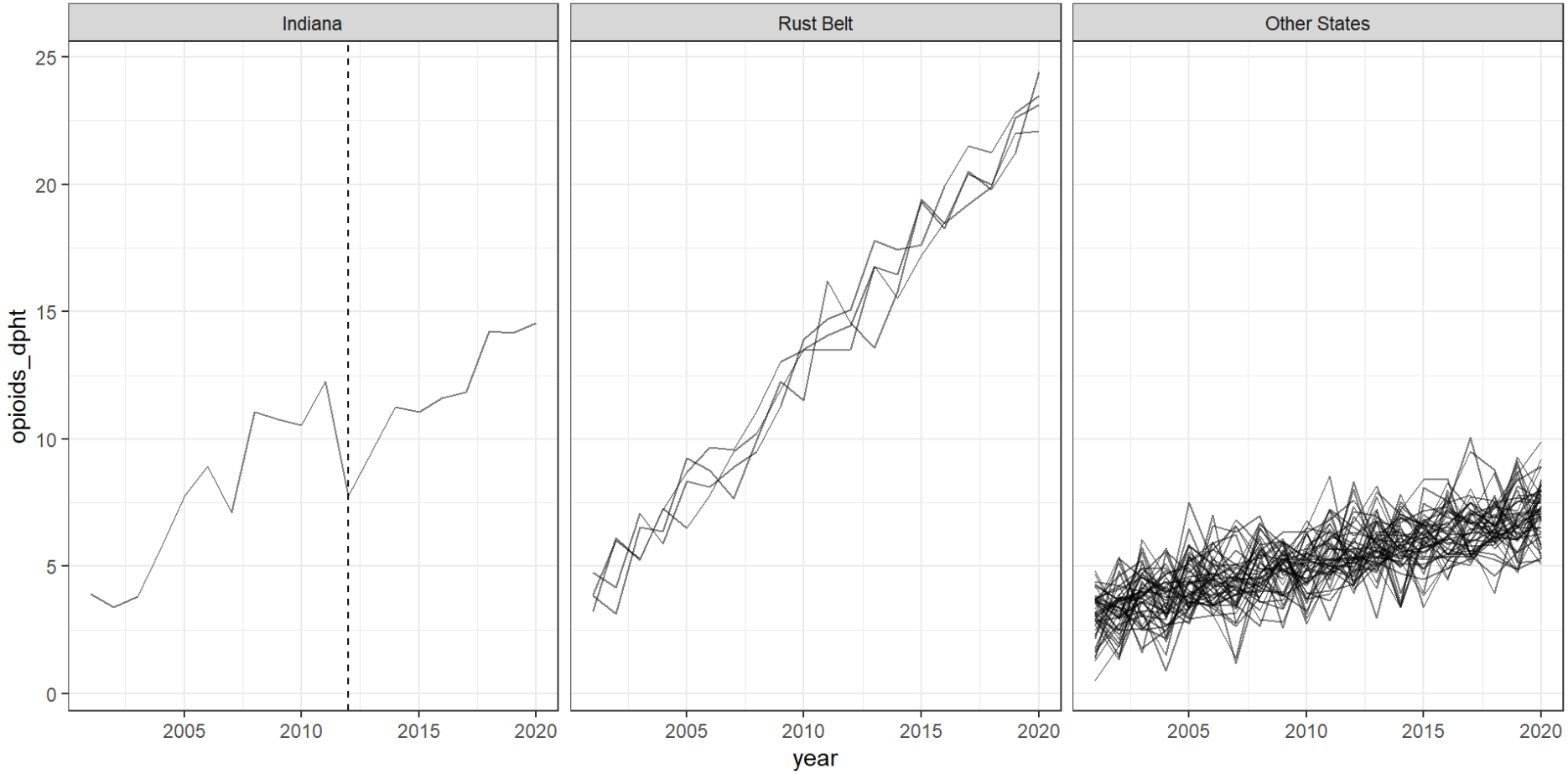
Rust Belt States Were Most Strongly Affected

Generating Some Example Data

```
opioids_df <- expand.grid(
  state = state.abb,
  year = 2001:2020
) %>%
  rowwise() %>%
  mutate(state_type = case_when(
    state %in% c("MI", "OH", "WV", "PA") ~ "Rust Belt",
    state == "IN" ~ "Indiana",
    TRUE ~ "Other States"),
    state_type = factor(state_type, levels = c("Indiana", "Rust Belt", "Other States")),
    rx_restrict = ifelse(state == "IN" & year >= 2012, 1, 0),
    opioids_dpht = case_when(
      state_type == "Rust Belt" ~ 1 * (year - 2000) + rnorm(1, 3, 1),
      state_type == "Indiana" ~
        0.8 * (year - 2000) + rnorm(1, 3, 1) - 4*rx_restrict,
      TRUE ~ 0.2 * (year - 2000) + rnorm(1, 3, 1)
    ))
```

Implies $\tau_{IN} = -4$ deaths per hundred thousand per year

Visualizing the Data



Does a DiD Model Give Us a Good Estimate?

```
opioid_did <- feols(opioids_dpht ~ rx_restrict | state + year,
                    data = opioids_df)
```

	(1)
rx_restrict	1.37*
	(0.33)
Num.Obs.	1000
R2	0.745

+ p < 0.1, * p < 0.05

Diagnosing the Problems

There were **two problems** with using the DiD design here…

1. We only had a single state that adopted treatment

- Estimates will be noisy and inferences with a single treated unit will be difficult

2. We had a clear violation of the parallel trends assumption

- (A few) Rust Belt states had steeper upwards trends. (Many) other states had flatter upwards trends.
- The DiD estimator (incorrectly) attributed the growing post-treatment divergence between Indiana and the latter to the treatment effect

Blending Time Trends

Let’s address issue number two (non-parallel trends) first.

Suppose we use a mix of states that (together) somewhat approximate Indiana.

- Let’s use all 4 Rust Belts states and 4 non-Rust Belt near neighbors

In other words, can we balance the time trends to match Indiana’s?

A Second Attempt

```
state_mix <- filter(opioids_df, state %in% c("IN", "MI", "OH", "WV", "PA",
                                             "IL", "IA", "SD", "TN"))

mix_did <- feols(opioids_dpht ~ rx_restrict | state + year,
                 data = state_mix)
```

	All States	Similar States
rx_restrict	1.37*	−2.12
	(0.33)	(1.57)
Num.Obs.	1000	180
R2	0.745	0.831

+ p < 0.1, * p < 0.05

Closer to the true τ (though insignificant…)

The Synthetic Control (SC)

To implement this systematically, we'd ideally have…

- A more principled way to select comparison units
- A better way to conduct inference with a single treated unit

Synthetic Control Design - A research design that estimates dynamics effects in a single treated unit by comparing it's outcomes to outcomes of a weighted combination of untreated units (the "synthetic control")

- Outcome trajectory of the SC provides the unobserved counterfactual (absent treatment) for the treated

Better than DiD when $n^{Treat} \approx 1$ OR the PTA does not hold

Implementing the Synthetic Control

Data Requirements for the SC Design

What data do we need to get credible estimates from the SC design?

1. A treated unit and a sufficient number of untreated units indexed by $j \in \{1, \cdots, J\}$. For inference, want $J \geq 20$ (we'll discuss why).
2. A balanced panel with time periods indexed by $t \in \{1, \cdots, t^*, \cdots T\}$, where treatment occurs at t^*
3. Data on *outcomes* (Y_t^{Treat}, Y_{jt}) and *predictive covariates* ($\mathbf{X}^{Treat}$, $\mathbf{X}^{J}$)
4. A sufficient number of pre-treatment time periods ($T^{Pre} = t^* - 1$)
   - The more the better.
5. A sufficient number of post-treatment time periods ($T^{Post} = T - t^* + 1$)
   - Depends on how long TEs persist.

The Synthetic Control Setup

The synthetic control design recovers an *individual treatment effect*.

The Estimand:

$$\tau_t^{Treat} = Y_t^{Treat}(1) - Y_t^{Treat}(0) \quad \forall \quad t \geq t^*,$$

The right-hand quantity is unobserved. We estimate it using the outcomes of a weighed combination of untreated units.

The Estimator:

$$\hat{\tau}_t^{Treat} = Y_t^{Treat} - \hat{Y}_t^{Synth} = Y_t^{Treat} - \sum_{j=1}^{J} \hat{w}_j Y_{jt}$$

Choosing the Weights

The synthetic control can be expressed as a set of weights

$$\hat{W} = \{\hat{w}_1, \cdots, \hat{w}_J\} \quad \text{s.t.} \quad \hat{w}_j \geq 0 \quad \text{and} \quad \Sigma_{j=1}^{j} \hat{w}_j = 1$$

that minimize the MSPE for treated outcomes in the pre-treatment period:

$$\frac{1}{T^{Pre}} \sum_{t<t*} \left(Y_t^{Treat} - \left(\hat{w}_1 Y_{1t} + \hat{w}_2 Y_{2t} + \cdots + \hat{w}_J Y_{Jt} \right) \right)^2.$$

The idea: Construct a synthetic control whose outcome *most closely resembles* that of the treated unit pre-treatment

- We explicitly try to match pre-trends of treated and (synthetic) control

An Exemplary Synthetic Control Analysis

One of the first applications of the synthetic control design: Abadie, Diamond, and Hainmueller (2010)

The Research Question: How did the passage of a massive tax hike on cigarettes (CA Proposition 99) affect cigarette sales?

- Policy went into effect in 1989
- Data on CA and other states spans a 30 year time period (1970 - 2000)
- Pool of untreated states includes 38 states that did *not* have any cigarette tax increases during this time period

Pre-Treatment Cigarette Sales, CA vs. rest of US

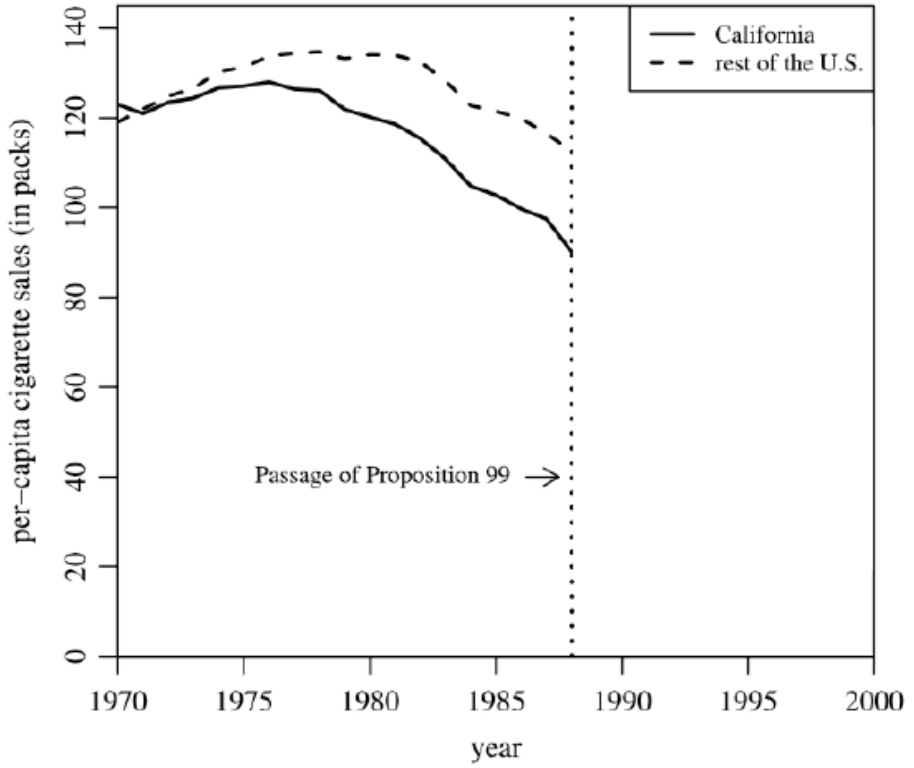


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Pre-Treatment Cigarette Sales, CA vs. the SC

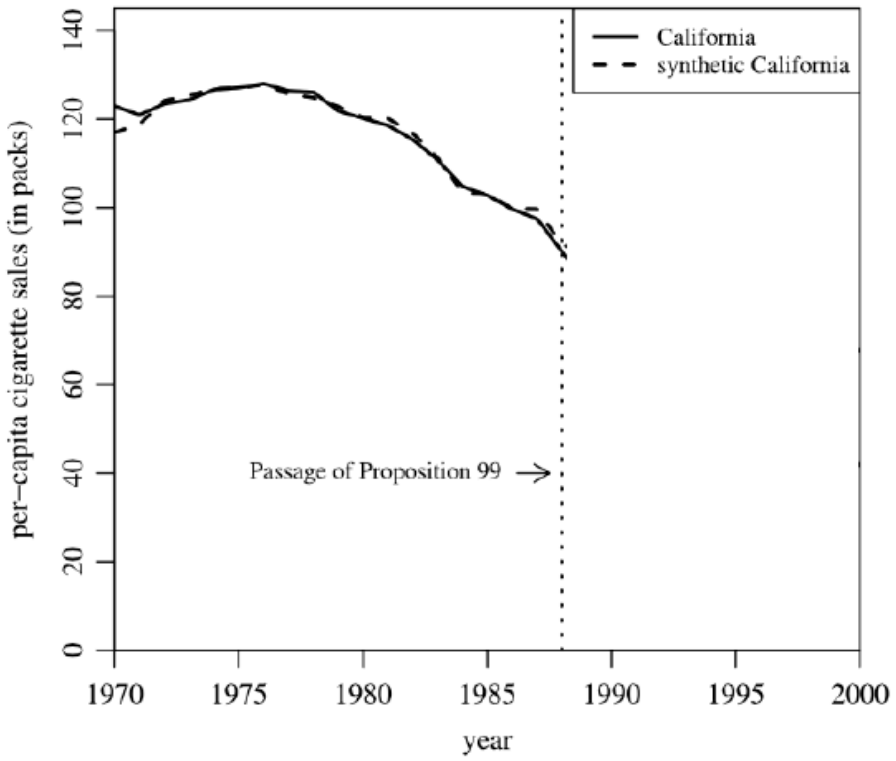


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

The Weights For Synthetic CA

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

Choosing the Covariates

Our weights are constructed using our matrices of covariates ($\mathbf{X}^{Treat}$, $\mathbf{X}^{J}$) to best approximate the pre-treatment outcome trend.

Which covariates should we include? Two common strategies:

- Additional variables that are predictive of the outcome
- Pre-treatment lagged versions of the outcome variable[1]

Typically variables are averaged across pre-treatment periods

- A sparser covariate matrix leads to sparser (more interpretable) weights

1. This strategy is similar to using lagged dependent variables in a regression.

Covariates for Constructing Synthetic 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.

Estimating Treatment Effects

With estimated weights, the next step is estimating the by-period treatment effect: $\hat{\tau}_t^{Treat} = Y_t^{Treat} - \hat{Y}_t^{Synth} = Y_t^{Treat} - \sum_{j=1}^{J} \hat{w}_j Y_{jt}$

Two ways to display the effects:

- Plot treated and SC outcomes (Y_t^{Treat}, $\hat{Y}_t^{Synth}$)
- Plot the treatment effect estimates ($\hat{\tau}_t^{Treat}$) themselves

A SC that fits the treated outcomes closely in the pre-period then diverges after is evidence of a treatment effect

- A close match in both periods is evidence against an effect
- A poor match in both periods is a poor estimate (not evidence of much)

Displaying Treated and SC Cigarette Sales

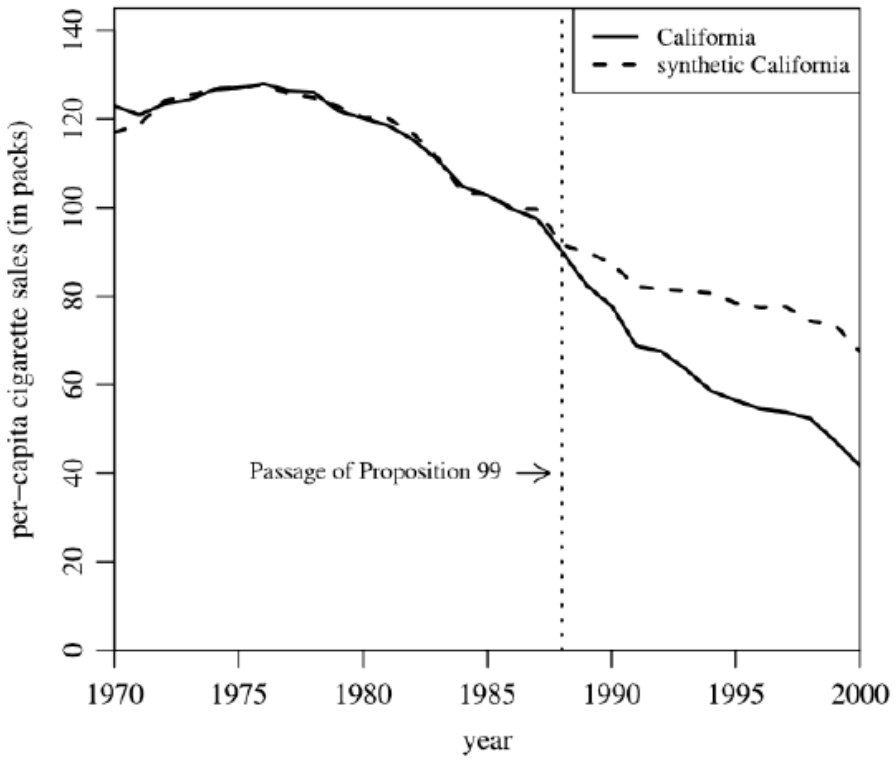
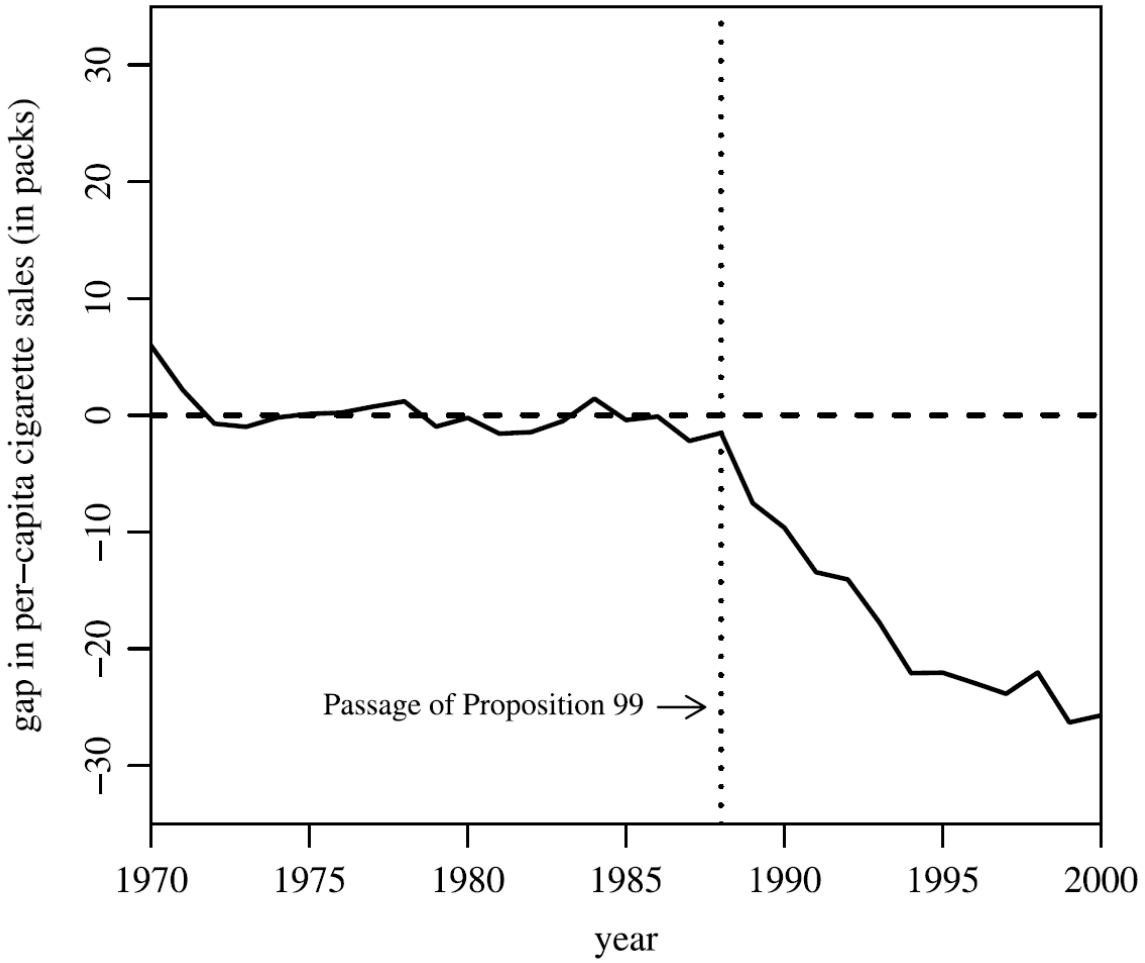


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Displaying Estimated TEs



Addressing Overfitting Concerns

One concern: A close fit in the pre-period may be due to overfitting

- As $J \to \infty$, can find a close pre-treatment fit regardless of whether you've approximated the underlying DGP
- **An implication** – Include units in the untreated pool judiciously (as similar as possible to the treated unit in observed and unobserved characteristics)

A solution from the ML lit: Create a training set ($t = 1, \cdots, t^{valid} - 1$) and a validation set ($t = t^{valid}, \cdots, t^* - 1$) of pre-treatment time periods

- Fit model on the former, examine predictive accuracy in the latter

Identifying Assumptions (1/2)

Identifying assumptions of the SC design:

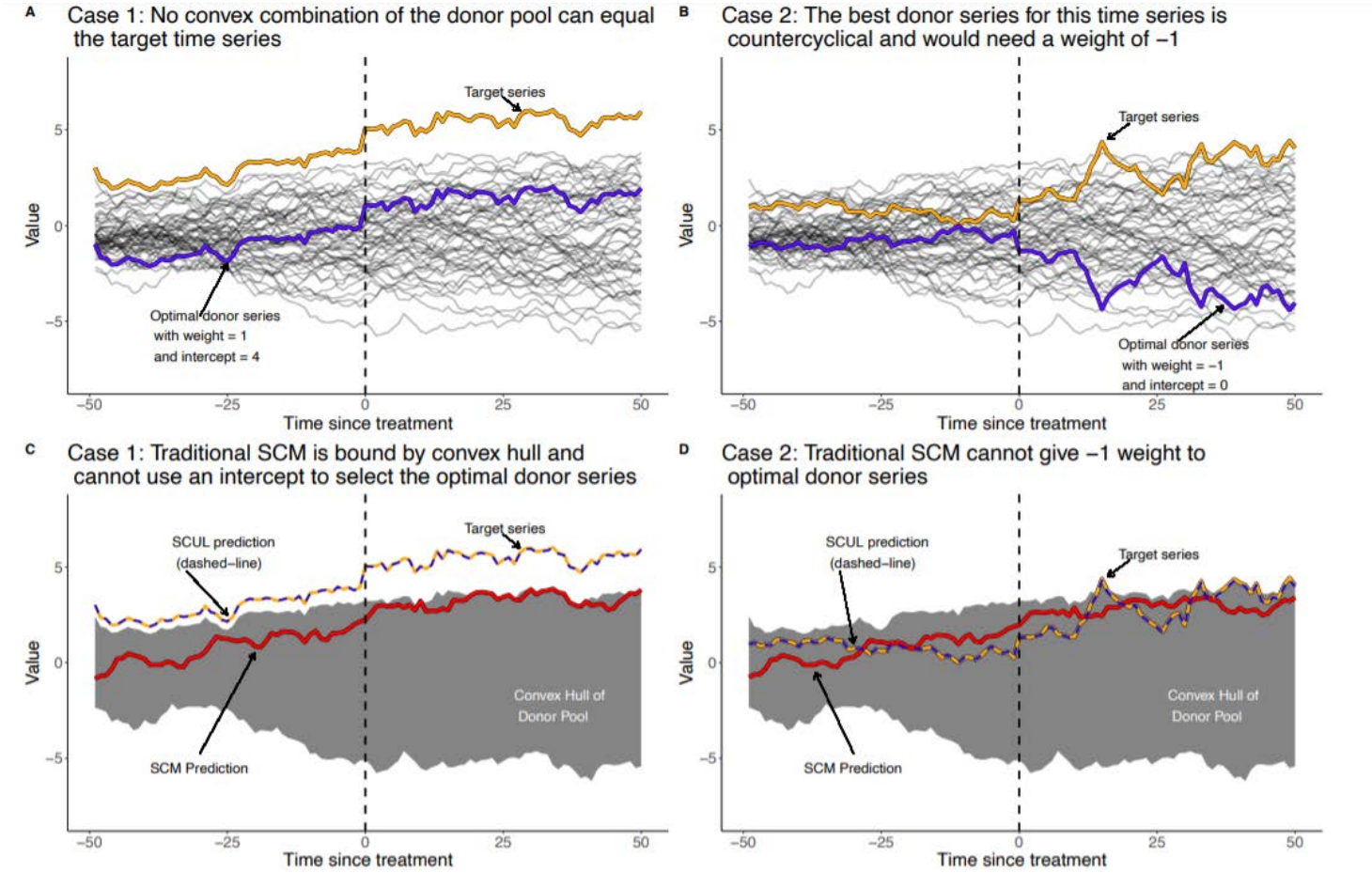
1. **Convex Hull Assumption** - A weighted combination of untreated units can approximate the POs of the treated unit:

$$Y_t^{Treat}(0) - w_1 Y_{1t}(0) - \cdots - w_J Y_{Jt} \approx 0$$

   Credible if a weighted combination of untreated units can approximate the pre-treatment outcomes of the treated unit:

$$Y_{Pre}^{Treat} - w_1 Y_{1,Pre} - \cdots - w_J Y_{J,Pre} \approx 0$$

Convex Combinations and Non-Negative Weights



Source: Hollingsworth and Wing (2020)

Identifying Assumptions (2/2)

The remaining assumptions are ones we've encountered already:

2. **No Anticipation** - Treated units do not change behavior before treatment comes into effect
3. **No Concurrent Unrelated Changes** - No other relevant developments at time t^* that affect solely the treated unit (no *bundled treatments*)
4. **No Interference (SUTVA)** - The treatment affects only the treated unit (no *spillovers*)

Inference for the Synthetic Control Design

A Randomization Inference Approach

How can we be confident that the effect we observed is statistically significant (i.e., not due to random noise)?

- **The solution:** A randomization inference approach

Randomization Inference - An approach to quantifying uncertainty in estimates that arises naturally from random assignment of treatments, rather than from hypothesized sampling from a large population.

- For more details on the general RI approach, see Athey and Imbens (2017)

Similar to bootstrapping (involves re-estimation of the treatment effect), but rather than randomly re-sampling units, you re-assign treatment

Testing the Sharp Null of No Effect

Under randomization inference, you test the sharp null hypothesis of no treatment effect for the treated unit in any time period

- $H_0 : \tau_t = 0 \quad \forall \quad t$
- $H_A : \tau_t \neq 0$

Procedure:

- Apply the SC method to each untreated unit *as if* it were treated
  - Put the treated unit in the pool with the (other) untreated units
- Calculate the TE estimate for each unit (treated and untreated)
- Compare the actual treated unit's effect estimate to the distribution of placebo effects

Randomization Inference for CA's Prop. 99 (1/2)

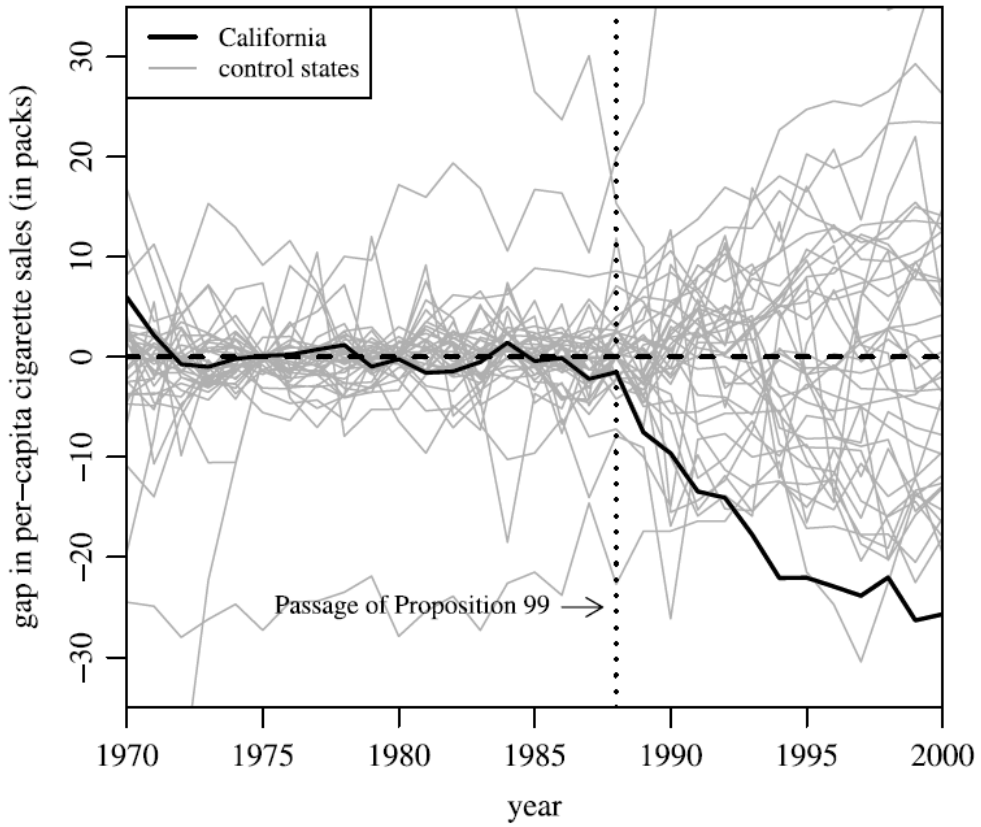


Figure 4. Per-capita cigarette sales gaps in California and placebo gaps in all 38 control states.

Randomization Inference for CA's Prop. 99 (2/2)

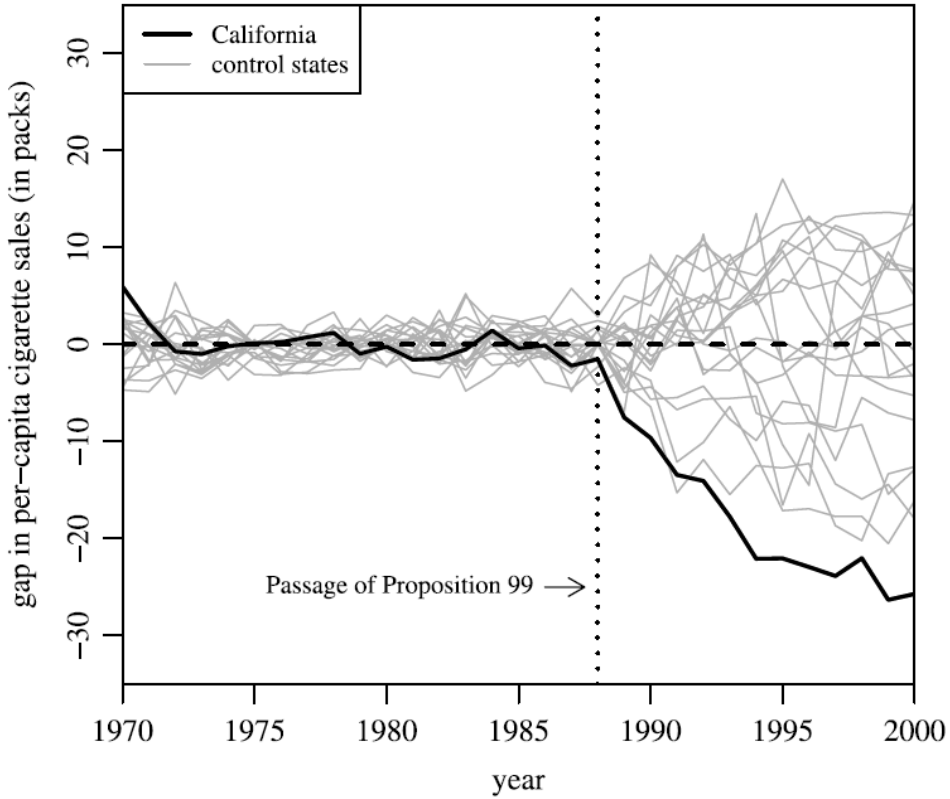


Figure 7. Per-capita cigarette sales gaps in California and placebo gaps in 19 control states (discards states with pre-Proposition 99 MSPE two times higher than California's).

Obtaining the p-value

Can obtain a p-value by comparing the ratio of the Post-/Pre- prediction error of the treated versus placebo tests

1. For treatment and all placebo SC fits, calculate

$$\text{RMSPE}_{Pre} = \sqrt{\frac{1}{t^* - 1}\sum_{t=1}^{t^*-1}\left(Y_t^{\text{“}Treat\text{”}} - \sum_{j=1}^{J} w_j Y_{jt}\right)^2}$$

2. Calculate RMSPE_{Post} similarly (for periods t^* to T)
3. Compute the ratio of $\frac{RMSPE_{Post}}{RMSPE_{Pre}}$
4. The p-value is the rank of the ratio for the treated divided by the total number of units (e.g., if largest out of 21, $p \approx 0.048 < 0.05$)

The p-value for CA's Proposition 99

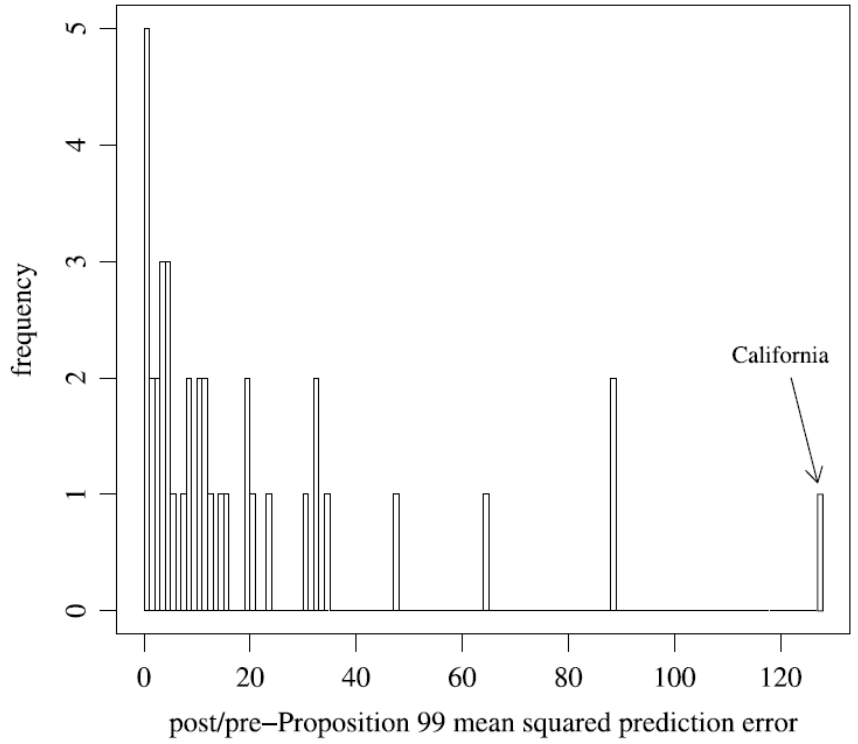


Figure 8. Ratio of post-Proposition 99 MSPE and pre-Proposition 99 MSPE: California and 38 control states.

p-value = 1/39 = 0.026

Advantages of Randomization Inference

This approach is particularly valuable for synthetic control designs because:

- Doesn’t require large sample approximations (important since SC often involves few units)
- Nonparametric, doesn’t rely on distributional assumptions
- Directly addresses the fundamental question: How likely is it to observe an effect this large by chance alone?

Limitations:

- Doesn’t provide sampling based inferential quantities, like CIs
- Doesn’t tell us how likely the effect is to generalize to other cases (not really an inference issue, rather a design issue with the traditional SC)

Synthetic Control with Multiple Treated Units

Estimation and Inference with $n^{Treat} > 1$

Can the SC design be used with more than 1 treated unit? Yes!

The simplest solution: Apply the methodology described above, and aggregate individual treatment effects at the end.

- Estimation and aggregation is straightforward (but perhaps tedious). The per-period SATE simple equals $\frac{1}{n^{Treat}} \sum_{i=1}^{n^{Treat}} \hat{\tau}_t$.
- Inference is more challenging. Requires a modified version of the permutation comparing the rank of the n^{treat} actual treatment effects ($\hat{\tau}_i$) to the J placebo effects.
- See Abadie and L'Hour (2021) for details on inference.

Generalized Synthetic Control and the Factor Model

An Alternative: The *generalized synthetic control* method (Xu 2017)

Builds off an intuition from Abadie et. al. (2010). One justification for "why" SC works: can capture an underlying *latent factor model*:

$$Y_{it}(0) = \beta_t \mathbf{X}_i + \underbrace{\mathbf{f}_t}_{\text{Factors}} \quad \underbrace{\lambda_i}_{\text{Factor Loadings}} + \epsilon_{it}$$

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

If we can estimate the model for $Y_{it}(0)$, we can estimate τ_{it}

Abadie et. al. (2010) prove that SC bounds the bias as a function of T^{Pre}:

$$\hat{Y}_{it}(0) \equiv \Sigma_{j=1}^{J} \hat{w}_j Y_{jt} \rightarrow Y_{it}(0) \quad \text{as} \quad T^{Pre} \rightarrow \infty$$

Understanding the Factor Model

What are the factors $\mathbf{f}_t$? Represent *time trends* or *time shocks*.

What are the factor loadings λ_i? Represent *unit exposure to time trends/shocks*.

The interaction of the two ($\mathbf{f}_t \times \lambda_i$) covers many situations.

- **Time fixed effects** – Shared exposure to time shocks: $\lambda_i = 1$
- **Unit fixed effects** – Time-invariant differences between units: $\mathbf{f}_t = 1$
- **Unit-specific trends** – Differential exposure to a common trend (linear, quadratic, etc.): $\mathbf{f}_t \times \lambda_i = \lambda_i(t + t^2 + \cdots)$

In the Opioid example, $\lambda_i = 1$ for Rust Belt states, $\lambda_i = 0.8$ for Indiana, $\lambda_i = 0.2$ for other states, and $\mathbf{f}_t = t - 2000$

Recovering the Factor Model

If we knew the factors and factor loadings, estimating $Y_{it}(0)$ (and thus τ_{it}) would be trivial.

```
opioids_df <- mutate(opioids_df,
                     factor_loadings = case_when(
                       state_type == "Rust Belt" ~ 1.0,
                       state_type == "Indiana" ~ 0.8,
                       state_type == "Other States" ~ 0.2))

factor_model <- feols(opioids_dpht ~ rx_restrict + I(factor_loadings * (year - 2000)),
                      data = opioids_df)

summary(factor_model)
```

```
OLS estimation, Dep. Var.: opioids_dpht
Observations: 1,000
Standard-errors: IID
                                 Estimate Std. Error  t value  Pr(>|t|)
(Intercept)                      3.086181   0.044917  68.7093 < 2.2e-16 ***
rx_restrict                     -4.006668   0.364569 -10.9901 < 2.2e-16 ***
I(factor_loadings * (year - 2000))  0.992378   0.010842  91.5301 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 1.03803   Adj. R2: 0.896615
```

An Empirical Example of Factors and Loadings

A research question: How does Election Day Registration (EDR) affect turnout?

- *The outcome (Y_{it}):* turnout in state i and election $t \in \{1920, \cdots, 2012\}$
- *Treated units:* states that implement EDR
- *Counterfactual:* (per-election) turnout without EDR
  - Estimated using Y_{it} in non-treated states

One plausible time shock: Removal of Jim Crow laws following the passage of the 1965 Voting Rights Act

- Should primarily affect Southern states

Recovering the Factor Model Using GSC

Xu (2017) proposes a generalized SC method that recovers factors + loadings

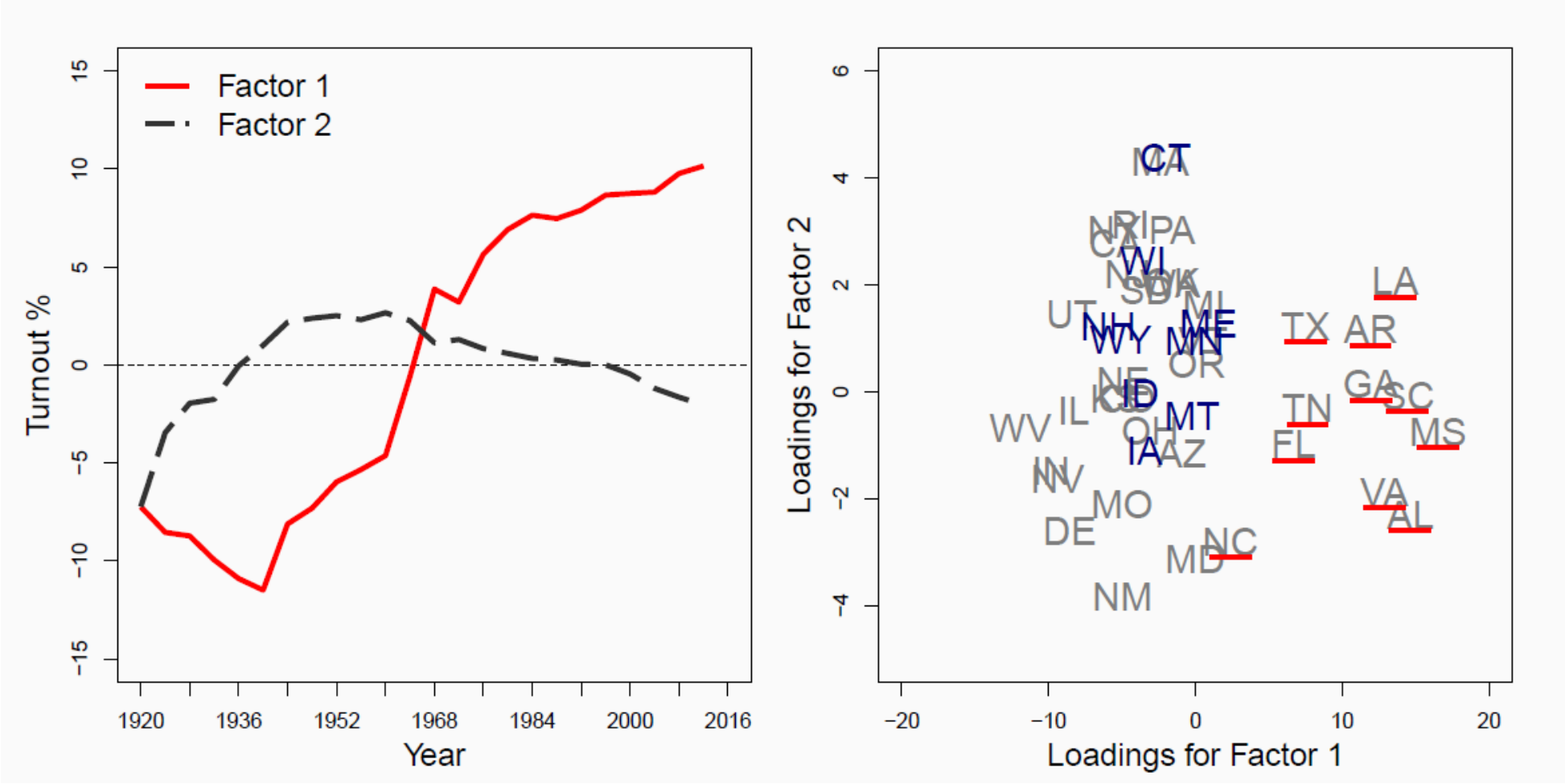
The Estimation Procedure:

1. Estimate β_t, $\mathbf{F}_t$, and $\lambda_i^{Untreated}$ using an Expectation-Maximization (EM) algorithm applied to untreated units
    - Choose the number *r* of factors + loadings using a Leave-One-Out (LOO) cross-validation procedure. Assume factors are orthogonal to each other.
2. Estimate loadings for treated (λ_i^{Treat}) that minimize pre-treatment MSPE
3. Calculate post-treatment counterfactuals and TEs using estimates:

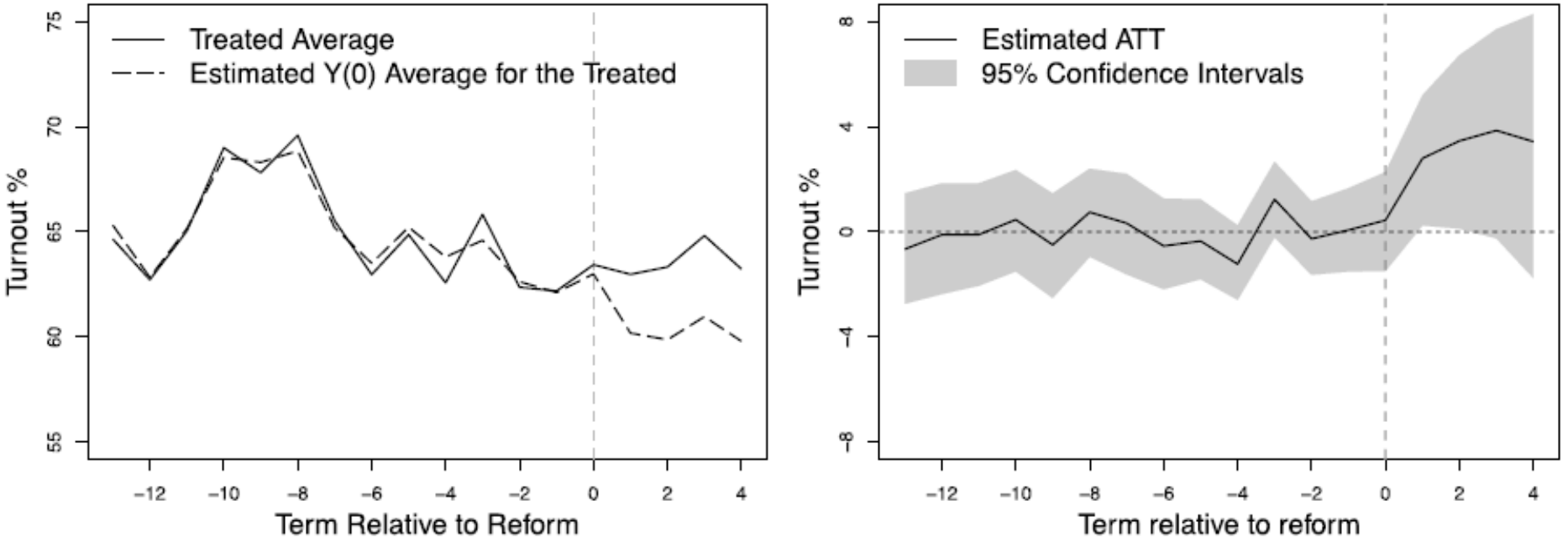
$$\hat{Y}_{it}^{Treat}(0) = \hat{\beta}_t \mathbf{X}_i^{Treat} + \hat{\mathbf{F}}_t \hat{\lambda}_i^{Treat}$$

$$\hat{\tau_{it}} = Y_{it} - \hat{Y}_{it}^{Treat}(0)$$

Factors and Loadings for the EDR Analysis



Estimated Counterfactuals and Effects



(b) Generalized synthetic control

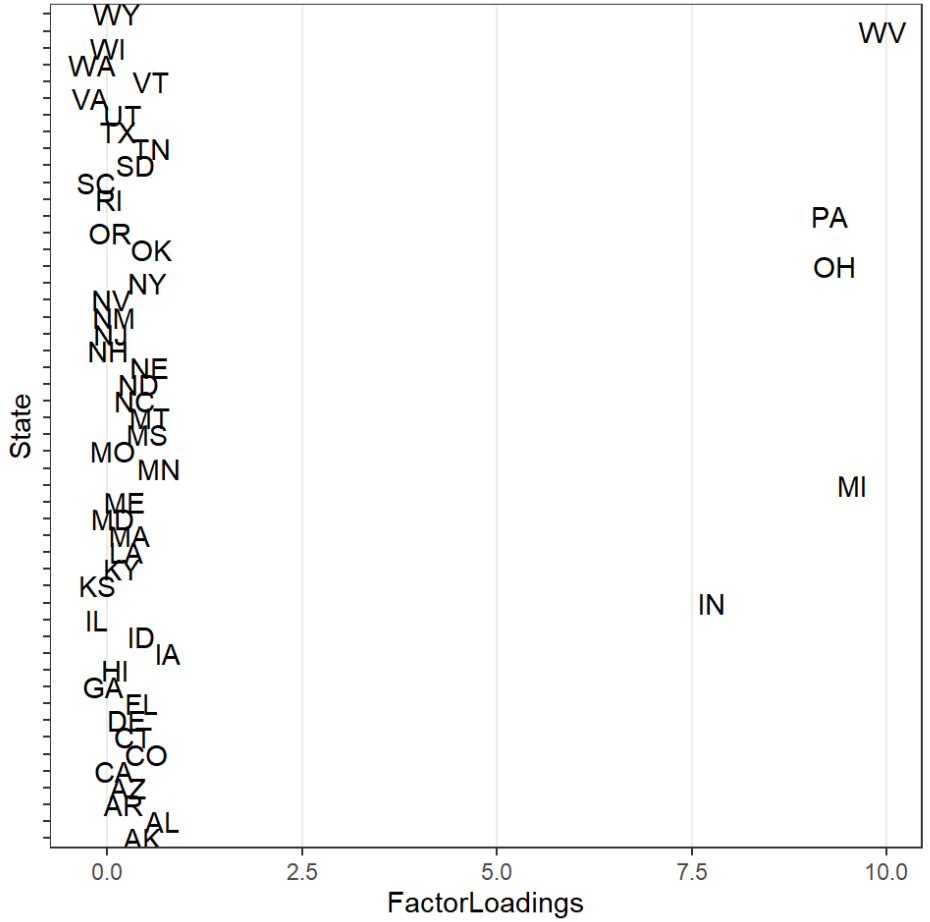
Figure 2. The effect of EDR on turnout: Main results.

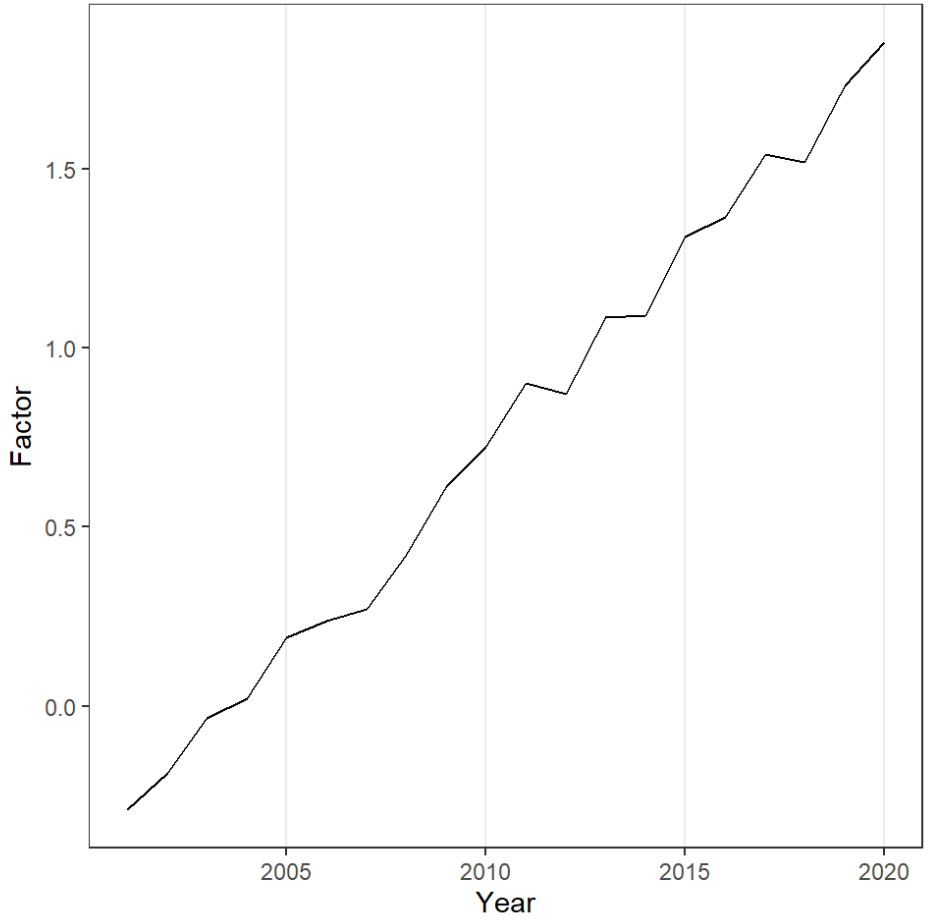
Using gsynth With the IN Example

```
model_gsynth <- gsynth(opioids_dpht ~ rx_restrict,
       data = opioids_df,
       index = c("state", "year"),
       force = "none",
       r = 1)

model_gsynth
```

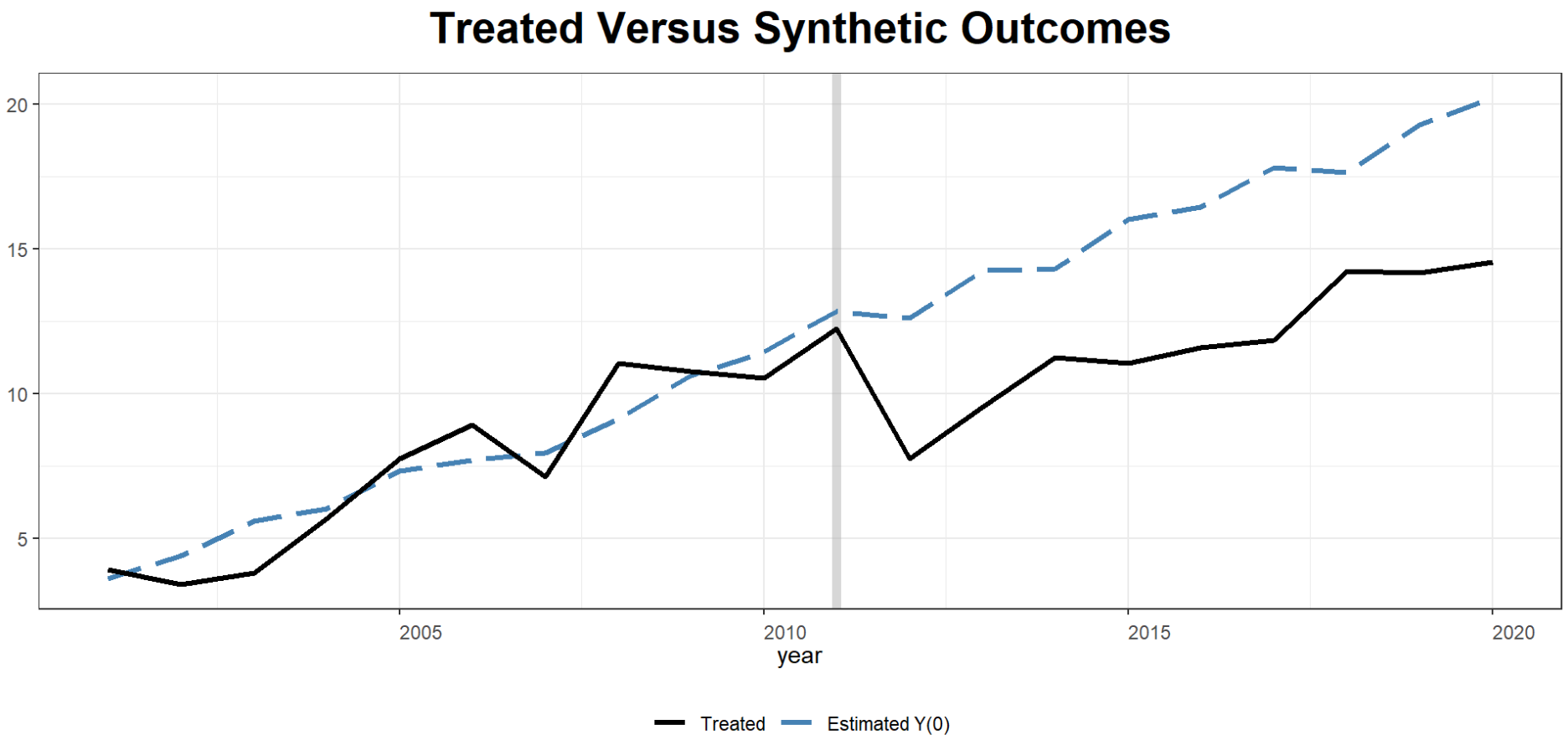
```
Call:
gsynth.formula(formula = opioids_dpht ~ rx_restrict, data = opioids_df,
    index = c("state", "year"), force = "none", r = 1, CV = FALSE)

Average Treatment Effect on the Treated:
[1] -4.736

   ~ by Period (including Pre-treatment Periods):
   2001    2002    2003    2004    2005    2006    2007    2008    2009    2010
 0.3164 -1.0016 -1.7795 -0.3423  0.4457  1.2132 -0.8241  1.9466  0.1609 -0.9328
   2011    2012    2013    2014    2015    2016    2017    2018    2019    2020
-0.5987 -4.8720 -4.7373 -3.0653 -4.9507 -4.8409 -5.9638 -3.3891 -5.1237 -5.6780

Uncertainty estimates not available.
```

Factors and Loadings for the IN Example

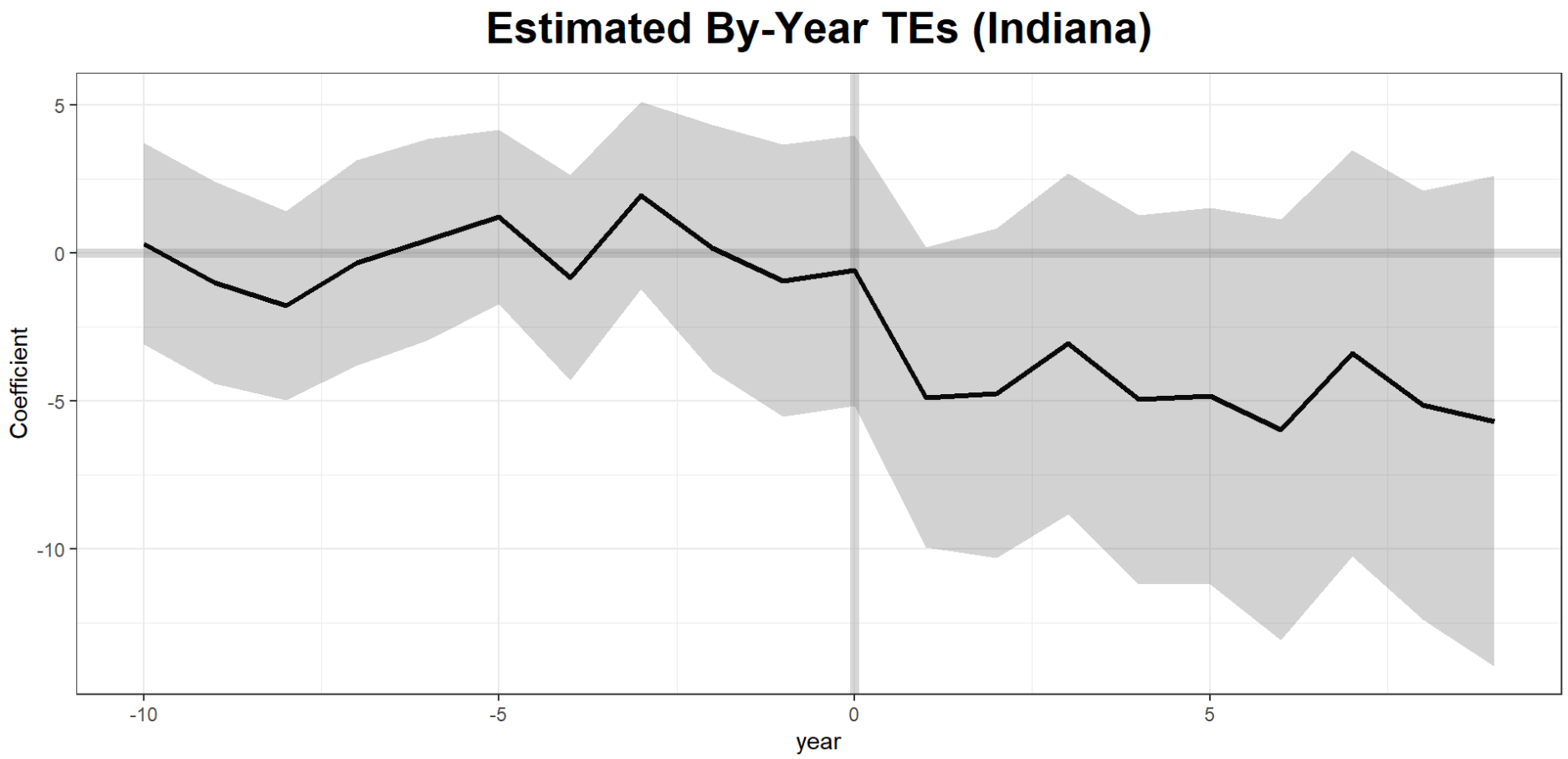




Plotting IN Outcomes vs. the Synthetic Control



Plotting the Per-Year Treatment Effects



Advantages and Limitations of the Generalized SC

The Generalized SC Method has some **attractive features:**

1. Easily extends SC to multiple treated units and staggered treatment
2. Can use information from *all* untreated units to model counterfactuals
3. Does not rely on a convex hull assumption
4. Produces sampling-centric inferences (e.g., CIs) for the ATT

Limitations:

1. Requires more data (especially T^{Pre}) than traditional SC
2. Relies on a Conditional Independence Assumption (CIA)
    - Conditional on covariates, factors, and loadings, no omitted confounders

Conducting Synthetic Control Analyses in R

To estimate a traditional synthetic control design, can use the `Synth` package

- Developed by Abadie, Diamond, and Hainmueller

To estimate a generalized synthetic control design, can use the `gsynth` package

- Developed by Yiquing Xu (creator of `panelView` and `interflex` as well!)

Let's see how to use these packages in our data lab…