14.661 Recitation 2: DD, SC

Andrea Manera

September 23, 2021

Potential Outcome: the What and How of Causality

- X_i: treatment actually administered to a unit i.

Potential Outcome: the What and How of Causality

- X_i: treatment actually administered to a unit i.
  - $X_i = 1$ administered, $\forall i \in$ T, treatment;
  - $X_i = 0$, $\forall i \in$ C, control.

Potential Outcome: the What and How of Causality

- X_i: treatment actually administered to a unit i.
  - $X_i = 1$ administered, $\forall i \in \mathrm{T}$, treatment;
  - $X_i = 0$, $\forall i \in \mathrm{C}$, control.
- Y_{ij} the outcomes of some unit i, after receiving treatment j
- $T_i = (Y_i | X_i = 1) - (Y_i | X_i = 0)$: treatment effect for unit i
- Two objects of interest:

$$ATT \equiv E_{i \in \mathrm{T}} [Y_{i1} - Y_{i0}]$$

Potential Outcome: the What and How of Causality

- X_i: treatment actually administered to a unit i.
  - $X_i = 1$ administered, $\forall i \in \mathrm{T}$, treatment;
  - $X_i = 0$, $\forall i \in \mathrm{C}$, control.
- Y_{ij} the outcomes of some unit i, after receiving treatment j
- $T_i = (Y_i|X_i = 1) - (Y_i|X_i = 0)$: treatment effect for unit i
- Two objects of interest:

$$ATT \equiv E_{i\in \mathrm{T}}\left[Y_{i1} - Y_{i0}\right]$$

$$\begin{aligned} ATE &\equiv E_{i\in \mathrm{pop}}\left[Y_{i1} - Y_{i0}\right] \\ &= Pr\left\{i \in \mathrm{C}\right\} E_{i\in \mathrm{C}}\left[Y_{i1} - Y_{i0}\right] + Pr\left\{i \in \mathrm{T}\right\} E_{i\in \mathrm{T}}\left[Y_{i1} - Y_{i0}\right] \end{aligned}$$

Potential Outcome: the What and How of Causality

- X_i: treatment actually administered to a unit i.
  - $X_i = 1$ administered, $\forall i \in \mathrm{T}$, treatment;
  - $X_i = 0$, $\forall i \in \mathrm{C}$, control.
- Y_{ij} the outcomes of some unit i, after receiving treatment j
- $T_i = (Y_i|X_i = 1) - (Y_i|X_i = 0)$: treatment effect for unit i
- Two objects of interest:

$$ATT \equiv E_{i\in\mathrm{T}}\left[Y_{i1} - Y_{i0}\right]$$

$$\begin{aligned} ATE &\equiv E_{i\in\mathrm{pop}}\left[Y_{i1} - Y_{i0}\right] \\ &= Pr\left\{i \in \mathrm{C}\right\} E_{i\in\mathrm{C}}\left[Y_{i1} - Y_{i0}\right] + Pr\left\{i \in \mathrm{T}\right\} E_{i\in\mathrm{T}}\left[Y_{i1} - Y_{i0}\right] \end{aligned}$$

- ATT: causal effect of intervention on treated units
- ATE: causal effect of intervention if scaled up to both treatment and control

Selection Bias

- Almost all papers estimate the ATT as:

$$\hat{ATT} = E_{i\in \mathrm{T}}[Y_{i1}] - E_{i\in \mathrm{C}}[Y_{i0}]$$
$$= E_{i\in \mathrm{T}}[Y_{i1}] \pm E_{i\in \mathrm{T}}[Y_{i0}] - E_{i\in \mathrm{C}}[Y_{i0}]$$
$$= ATT + E_{i\in \mathrm{T}}[Y_{i0}] - E_{i\in \mathrm{C}}[Y_{i0}]$$

Selection Bias

- Almost all papers estimate the ATT as:

$$\begin{aligned}
A\hat{T}T &= E_{i\in \mathrm{T}}\left[Y_{i1}\right] - E_{i\in \mathrm{C}}\left[Y_{i0}\right] \\
&= E_{i\in \mathrm{T}}\left[Y_{i1}\right] \pm E_{i\in \mathrm{T}}\left[Y_{i0}\right] - E_{i\in \mathrm{C}}\left[Y_{i0}\right] \\
&= ATT + E_{i\in \mathrm{T}}\left[Y_{i0}\right] - E_{i\in \mathrm{C}}\left[Y_{i0}\right]
\end{aligned}$$

- The rightmost term is selection bias, difference between treatment and control *in the absence of treatment* (counterfactual!)
- $E_{i\in \mathrm{T}}\left[Y_{i0}\right] - E_{i\in \mathrm{C}}\left[Y_{i0}\right]$ is treatment and control balance in the counterfactual world where T are not treated

Selection Bias

- Almost all papers estimate the ATT as:

$$\begin{aligned} A\hat{T}T &= E_{i\in \mathrm{T}}\left[Y_{i1}\right] - E_{i\in \mathrm{C}}\left[Y_{i0}\right] \\ &= E_{i\in \mathrm{T}}\left[Y_{i1}\right] \pm E_{i\in \mathrm{T}}\left[Y_{i0}\right] - E_{i\in \mathrm{C}}\left[Y_{i0}\right] \\ &= ATT + E_{i\in \mathrm{T}}\left[Y_{i0}\right] - E_{i\in \mathrm{C}}\left[Y_{i0}\right] \end{aligned}$$

- The rightmost term is selection bias, difference between treatment and control *in the absence of treatment* (counterfactual!)
- $E_{i\in \mathrm{T}}\left[Y_{i0}\right] - E_{i\in \mathrm{C}}\left[Y_{i0}\right]$ is treatment and control balance in the counterfactual world where T are not treated
- Literally *almost every empirical issue is about selection bias!*

Selection Bias

- Almost all papers estimate the ATT as:

$$\begin{aligned} A\hat{T}T &= E_{i\in \mathrm{T}}[Y_{i1}] - E_{i\in \mathrm{C}}[Y_{i0}] \\ &= E_{i\in \mathrm{T}}[Y_{i1}] \pm E_{i\in \mathrm{T}}[Y_{i0}] - E_{i\in \mathrm{C}}[Y_{i0}] \\ &= ATT + E_{i\in \mathrm{T}}[Y_{i0}] - E_{i\in \mathrm{C}}[Y_{i0}] \end{aligned}$$

- The rightmost term is selection bias, difference between treatment and control *in the absence of treatment* (counterfactual!)
- $E_{i\in \mathrm{T}}[Y_{i0}] - E_{i\in \mathrm{C}}[Y_{i0}]$ is treatment and control balance in the counterfactual world where T are not treated
- Literally *almost every empirical issue is about selection bias!*
- Caveat: Even if $A\hat{T}T = ATT$, it might be that $ATE \neq ATT$, since the latter requires:

$$E_{i\in \mathrm{T}}[Y_{i1}] - E_{i\in \mathrm{C}}[Y_{i1}]!$$

Example: Parallel Trends

- In Rubin's Notation diff-in-diff has:

$$
\begin{aligned}
\hat{ATT} &= E_{i\in \mathrm{T}}[\Delta Y_{i1}] - E_{i\in \mathrm{C}}[\Delta Y_{i0}] \\
&= ATT + \underbrace{E_{i\in \mathrm{T}}[\Delta Y_{i0}] - E_{i\in \mathrm{C}}[\Delta Y_{i0}]}_{\text{VIOLATION}}
\end{aligned}
$$

- Selection bias is now called "Parallel Trends"

Many States, Treated Variably

- Card (1992) makes the *federal min* into a DD experiment using an equation like

$$y_{ist} = \gamma_s + \lambda_t + \delta(fa_s \cdot d_t) + \varepsilon_{ist}, \tag{1}$$

where fa_s is *fraction affected* in each state (pre-increase proportion of teen labor force earning $< 3.80\$$) and d_t is a dummy for observations in 1990, after increase.

- Card (1992) used two periods, before and after and 51 states

Many States, Treated Variably

- Card (1992) makes the *federal min* into a DD experiment using an equation like

$$y_{ist} = \gamma_s + \lambda_t + \delta(\mathrm{fa}_s \cdot d_t) + \varepsilon_{ist}, \tag{1}$$

where fa_s is *fraction affected* in each state (pre-increase proportion of teen labor force earning $< 3.80\$$) and d_t is a dummy for observations in 1990, after increase.

- Card (1992) used two periods, before and after and 51 states
- Double-Diff: γ_s differences across states, λ_t across periods
- Two periods: levels w/fixed (state) effects = first differences:

$$\Delta\bar{y}_s = \lambda^* + \delta\mathrm{fa}_s + \Delta\bar{\varepsilon}_s, \tag{2}$$

where $\Delta\bar{y}_s$ is the change in teen employment in state s and $\Delta\bar{\varepsilon}_s$ is the differenced error

"Event Studies": Design and Diagnostics

$$y_{st} = \gamma_s + \lambda_t + \sum_{\tau=-T_{\text{pre}},\tau\neq 1}^{T_{\text{post}}} \delta_\tau \mathrm{d}_{s,t,\tau} + X'_{st}\beta + \varepsilon_{st},$$

$$\mathrm{d}_{s,t,\tau} \equiv 1\{s \text{ received treatment } \tau \text{ periods ago}\}$$

- If τ is negative, δ_τ gives the *pre-trend*, or *anticipatory effects*. If significant, trouble for parallel trends!
- Usually omit $\tau = -1$, normalize by period just before treatment.

"Event Studies": Design and Diagnostics

$$y_{st} = \gamma_s + \lambda_t + \sum_{\tau=-T_{\text{pre}},\tau\neq 1}^{T_{\text{post}}} \boxed{\delta_\tau} d_{s,t,\tau} + X'_{st}\beta + \varepsilon_{st},$$

$$d_{s,t,\tau} \equiv 1\{s \text{ received treatment } \tau \text{ periods ago}\}$$

- If τ is negative, δ_τ gives the *pre-trend*, or *anticipatory effects*. If significant, trouble for parallel trends!
- Usually omit $\tau = -1$, normalize by period just before treatment.
- Beware: if treatment period is not the same for all treated units, δ_τ, $\tau < 0$ are spurious
- Use Sun and Abraham (2020): Interaction-weighted estimator!

"Event Studies": Design and Diagnostics

$$y_{st} = \gamma_s + \lambda_t + \sum_{\tau=-T_{\text{pre}},\tau\neq 1}^{T_{\text{post}}} \delta_\tau d_{s,t,\tau} + X'_{st}\beta + \varepsilon_{st},$$

$$d_{s,t,\tau} \equiv 1\{s \text{ received treatment } \tau \text{ periods ago}\}$$

- If τ is negative, δ_τ gives the *pre-trend*, or *anticipatory effects*. If significant, trouble for parallel trends!
- Usually omit $\tau = -1$, normalize by period just before treatment.
- Beware: if treatment period is not the same for all treated units, δ_τ, $\tau < 0$ are spurious
- Use Sun and Abraham (2020): Interaction-weighted estimator!
- Unit/covariate time trends (3 periods min)

"Event Studies": Design and Diagnostics

$$y_{st} = \gamma_s + \lambda_t + \sum_{\tau=-T_{\text{pre}},\tau\neq 1}^{T_{\text{post}}} \delta_\tau d_{s,t,\tau} + X'_{st}\beta + \varepsilon_{st},$$

$$d_{s,t,\tau} \equiv 1\{s \text{ received treatment } \tau \text{ periods ago}\}$$

- If τ is negative, δ_τ gives the *pre-trend*, or *anticipatory effects*. If significant, trouble for parallel trends!
- Usually omit $\tau = -1$, normalize by period just before treatment.
- Beware: if treatment period is not the same for all treated units, δ_τ, $\tau < 0$ are spurious
- Use Sun and Abraham (2020): Interaction-weighted estimator!
- Unit/covariate time trends (3 periods min)
- Randomization/exact p-values

Nice Graphs!

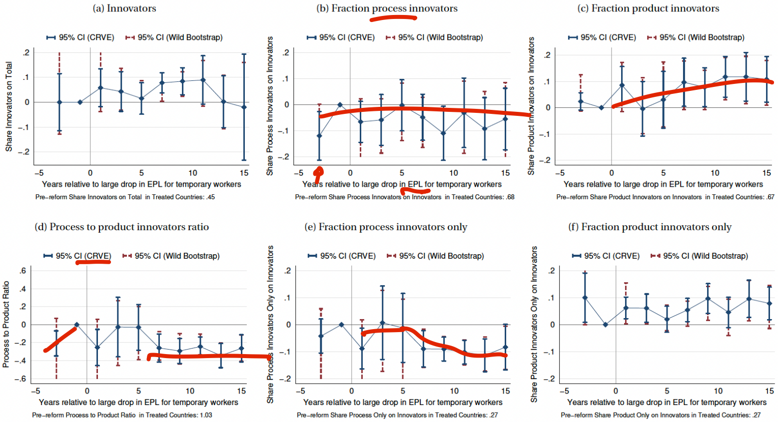
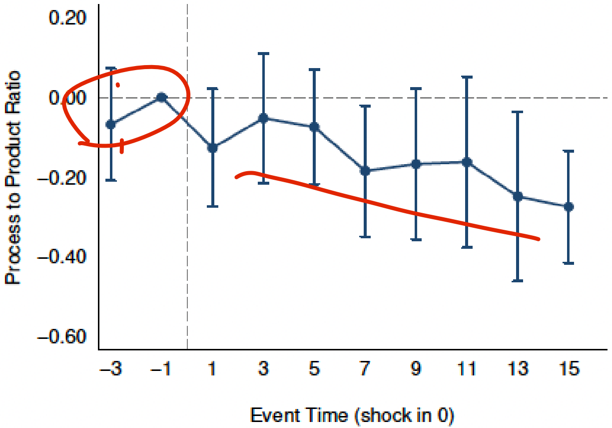


Figure 3: Main Results: effect of large EPL drop on innovators and product/process innovation

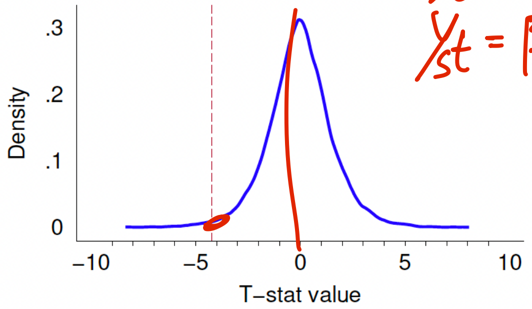
Interaction-Weighted

(f) IW estimator



Randomization

(c) Total randomization



T−stat est.: −4.225. Rand. p−value: 0.016.
Coeff. est.: −0.251. Rand. Avg.: −0.0014.

Figure 5: Process on product ratio: Permutation tests

$Y_t = \beta x_t + \gamma' \overline{X_t}$

$\tilde{Y}_{st} = \beta \tilde{x}_t + \tilde{\varepsilon}_t$

$\beta_i \quad i \in [1, 100$

Synthetic Control (Abadie et al., JEL 2021)

Abadie et al. (2003, 2010):

- Control group is called "donor pool"
- Covariates are called "predictors"

Synthetic Control (Abadie et al., JEL 2021)

Abadie et al. (2003, 2010):

- Control group is called "donor pool"
- Covariates are called "predictors"
- Synthetic controls use a weighted average of comparison units to match lagged predictors:
    - Idea: if you are similar on observables you are also on unobservables
    - Not necessarily, but often, matches also pre-treatment outcomes

Notation In Abadie Case

- The data consist of observations on regions i at time t for $i = 1, \ldots, J+1$, and $t = 1, \ldots, T$, where $1 \leq T_0 < T$ is the intervention date

Notation In Abadie Case

- The data consist of observations on regions i at time t for $i = 1, \ldots, J+1$, and $t = 1, \ldots, T$, where $1 \leq T_0 < T$ is the intervention date
- $\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$ is the effect of the intervention for unit i at time $t > T_0$. The first unit is treated and the aim is to estimate $(\alpha_{1T_0+1}, \ldots, \alpha_{1T})$. For $t > T_0$,

$$\alpha_{1t} = Y_{1t}^{I} - Y_{1t}^{N} = Y_{1t} - Y_{1t}^{N}.$$

Y_{1t}^{I} is observed in post-intervention periods. *Counterfactual* Y_{1t}^{N}:

$$Y_{1t}^{N} = \sum_{j=2}^{J+1} w_j Y_{jt}^{N},$$

Notation In Abadie Case

- The data consist of observations on regions i at time t for $i = 1, \ldots, J+1$, and $t = 1, \ldots, T$, where $1 \leq T_0 < T$ is the intervention date
- $\alpha_{it} = Y^I_{it} - Y^N_{it}$ is the effect of the intervention for unit i at time $t > T_0$. The first unit is treated and the aim is to estimate $(\alpha_{1T_0+1}, \ldots, \alpha_{1T})$. For $t > T_0$,

$$\alpha_{1t} = Y^I_{1t} - Y^N_{1t} = Y_{1t} - Y^N_{1t}.$$

  Y^I_{1t} is observed in post-intervention periods. *Counterfactual* Y^N_{1t}:

$$Y^N_{1t} = \sum_{j=2}^{J+1} w_j Y^N_{jt},$$

- Weighted average of untreated units, effectively a vastly more general DD!

Choice of SC

The *Synthetic Control* is defined as the vector of weights $\boldsymbol{W}$. How to choose it?

Choice of SC

The *Synthetic Control* is defined as the vector of weights $\boldsymbol{W}$. How to choose it?

Abadie, Diamond, and Hainmuller (2010):

- Choose a set of covariates ("predictors" for the dep. variable), $h \in \mathcal{H}$.
- Choose a set of importance weights v_h

Choice of SC

The *Synthetic Control* is defined as the vector of weights $\boldsymbol{W}$. How to choose it?

Abadie, Diamond, and Hainmuller (2010):

- Choose a set of covariates ("predictors" for the dep. variable), $h \in \mathcal{H}$.
- Choose a set of importance weights v_h
- Synthetic control solves:

$$\boldsymbol{W}(\boldsymbol{V}) = \arg\min_{\boldsymbol{w} \geq 0} \left(\sum_{h \in \mathcal{H}} v_h \left[X_{1,h} - \sum_{j=2}^{J+1} w_j X_{j,h} \right]^2 \right)^{\frac{1}{2}}$$

Choice of SC

The *Synthetic Control* is defined as the vector of weights $\boldsymbol{W}$. How to choose it?

Abadie, Diamond, and Hainmuller (2010):

- Choose a set of covariates ("predictors" for the dep. variable), $h \in \mathcal{H}$.
- Choose a set of importance weights v_h
- Synthetic control solves:

$$\boldsymbol{W}(\boldsymbol{V}) = \arg\min_{\boldsymbol{w} \geq 0} \left(\sum_{h \in \mathcal{H}} v_h \left[X_{1,h} - \sum_{j=2}^{J+1} w_j X_{j,h} \right]^2 \right)^{\frac{1}{2}}$$

- But how to choose v_h?

Choosing weights

- Choose $\boldsymbol{V}$ such that:

$$\boldsymbol{V} = \arg\min_{v} \text{MSPE}(v) \equiv \sum_{t<T_0} \underbrace{\left(Y_{1t} - \sum_{j=2}^{J+1} w_j(v) Y_{jt} \right)^2}$$

Choosing weights

- Choose $\boldsymbol{V}$ such that:

$$\boldsymbol{V} = \arg\min_{\boldsymbol{v}} \text{MSPE}(\boldsymbol{v}) \equiv \sum_{t<T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(\boldsymbol{v}) Y_{jt} \right)^2$$

- Minimizes the mean square prediction error (MSPE) of outcome in the pre-period.
- A more sophisticated way is using *out-of-sample validation* (see JEL and Abadie et al, 2015).

Choosing weights

- Choose $\boldsymbol{V}$ such that:

$$\boldsymbol{V} = \arg\min_{\boldsymbol{v}} \mathrm{MSPE}(\boldsymbol{v}) \equiv \sum_{t<T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(\boldsymbol{v}) Y_{jt} \right)^2$$

- Minimizes the mean square prediction error (MSPE) of outcome in the pre-period.
- A more sophisticated way is using *out-of-sample validation* (see JEL and Abadie et al, 2015).
- Final result:
  - Synthetic control is a set of nonnegative weights, which you can report in a table

Choosing weights

- Choose $\boldsymbol{V}$ such that:

$$\boldsymbol{V} = \arg\min_{\boldsymbol{v}} \text{MSPE}(\boldsymbol{v}) \equiv \sum_{t<T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(\boldsymbol{v}) Y_{jt} \right)^2$$

- Minimizes the mean square prediction error (MSPE) of outcome in the pre-period.
- A more sophisticated way is using *out-of-sample validation* (see JEL and Abadie et al, 2015).
- Final result:
  - Synthetic control is a set of nonnegative weights, which you can report in a table
  - Also regression creates weights for each observation, but nobody sees them! (Cunningham, 2021)

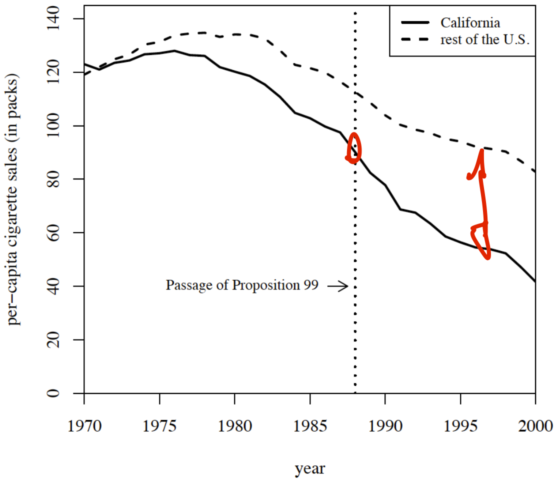
Choosing weights

- Choose $\boldsymbol{V}$ such that:

$$\boldsymbol{V} = \arg\min_{\boldsymbol{v}} \text{MSPE}(\boldsymbol{v}) \equiv \sum_{t<T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(\boldsymbol{v}) Y_{jt} \right)^2$$

- Minimizes the mean square prediction error (MSPE) of outcome in the pre-period.
- A more sophisticated way is using *out-of-sample validation* (see JEL and Abadie et al, 2015).
- Final result:
  - Synthetic control is a set of nonnegative weights, which you can report in a table
  - Also regression creates weights for each observation, but nobody sees them! (Cunningham, 2021)
  - Estimator bias decreases with length of pre-period.

All Data


California
rest of the U.S.
Passage of Proposition 99
per−capita cigarette sales (in packs)
0
20
40
60
80
100
120
140
1970
1975
1980
1985
1990
1995
2000
year

Matching Table

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

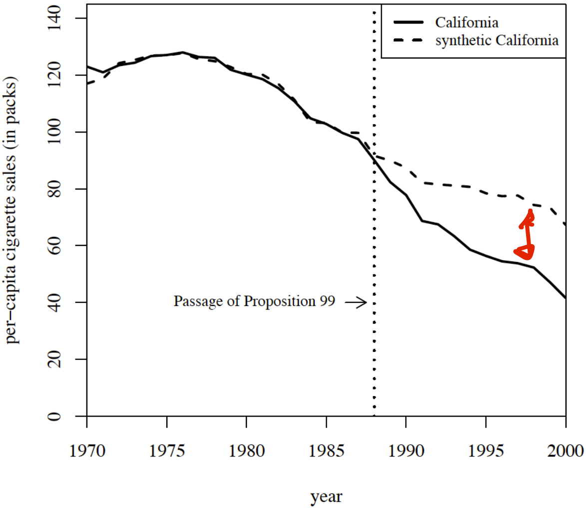
Note: All variables except lagged cigarette sales are averaged for the 1980-1988 period (beer consumption is averaged 1984-1988). Cigarette sales are measured in packs.

Matching Weights (?)

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

Result


California
synthetic California
Passage of Proposition 99
per–capita cigarette sales (in packs)
0
20
40
60
80
100
120
140
1970
1975
1980
1985
1990
1995
2000
year

What About Inference? Randomize!

- Empirical CDF of *real* treatment post-outcome relative to *random* treatment ("*exact p-value*")

What About Inference? Randomize!

- Empirical CDF of *real* treatment post-outcome relative to *random* treatment ("*exact p-value*")
- Abadie et al. (2010) compute a synthetic control W_j for each unit in the sample

What About Inference? Randomize!

- Empirical CDF of *real* treatment post-outcome relative to *random* treatment ("*exact p-value*")
- Abadie et al. (2010) compute a synthetic control W_j for each unit in the sample
- Then compute:

$$\underline{\text{RMSPE}_{j,\boxed{t_1,t_2}}} = \left(\frac{1}{t_2 - t_1 + 1} \sum_{\underline{t \in [t_1, t_2]}} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(\upsilon) Y_{jt} \right)^2 \right)^{\frac{1}{2}}.$$

What About Inference? Randomize!

- Empirical CDF of *real* treatment post-outcome relative to *random* treatment ("*exact p-value*")
- Abadie et al. (2010) compute a synthetic control W_j for each unit in the sample
- Then compute:

$$\text{RMSPE}_{j,t_1,t_2} = \left(\frac{1}{t_2 - t_1 + 1} \sum_{t \in [t_1, t_2]} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(v) Y_{jt} \right)^2 \right)^{\frac{1}{2}} .$$

- Compute the ratio of RMSPE post versus pre:

$$r_j = \frac{\text{RMSPE}_{j,T_0+1,T}}{\text{RMSPE}_{j,1,T_0}}$$

What About Inference? Randomize!

- Empirical CDF of *real* treatment post-outcome relative to *random* treatment ("*exact p-value*")
- Abadie et al. (2010) compute a synthetic control W_j for each unit in the sample
- Then compute:

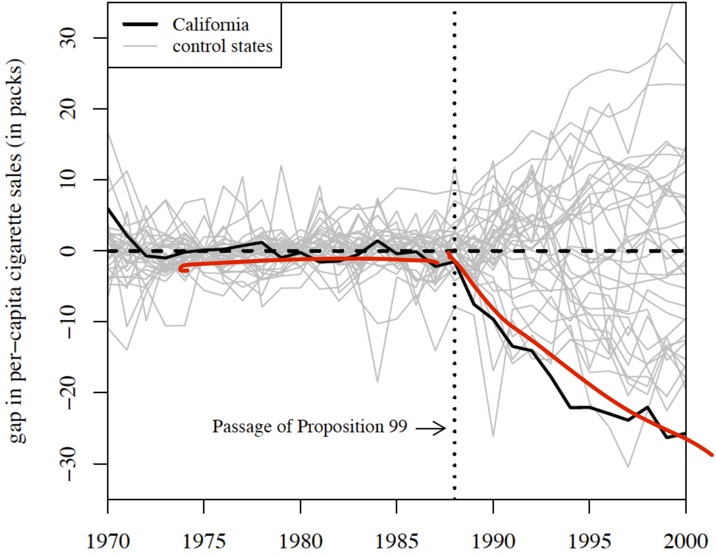
$$\text{RMSPE}_{j,t_1,t_2} = \left(\frac{1}{t_2 - t_1 + 1} \sum_{t \in [t_1, t_2]} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j(\upsilon) Y_{jt} \right)^2 \right)^{\frac{1}{2}}.$$

- Compute the ratio of RMSPE post versus pre:

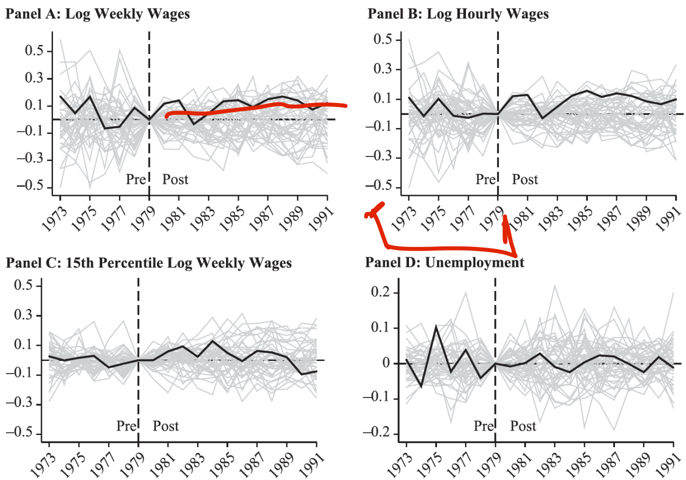
$$r_j = \boxed{\frac{\text{RMSPE}_{j,T_0+1,T}}{\text{RMSPE}_{j,1,T_0}}}$$

- Show the *permutation distribution* of r_j or compute the p-value as the empirical inverse CDF of r_j

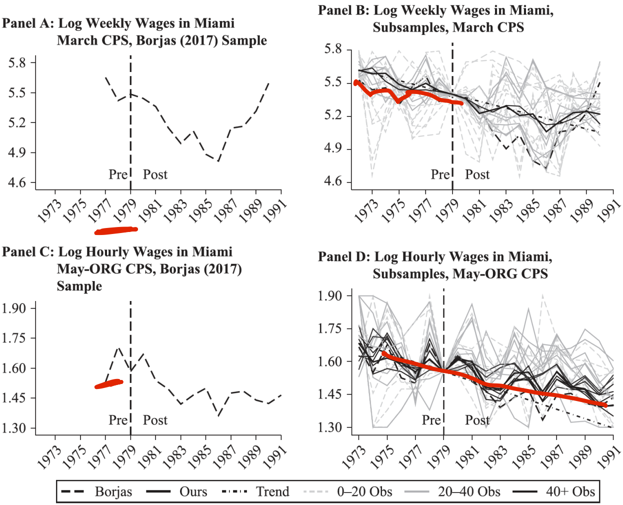
Show all Placebo Gaps, $Y_{j,t} - \hat{Y}^N_{j,t}$



Last Words on Mariel? Peri and Yasenov (2018)



Peri and Yasenov v. Borjas



Peri and Yasenov vs. Josh and Krueger