

# Canonical Research Designs II: Difference-in-Differences II: Event Studies, Synthetic Control, and Synthetic DiD

Paul Goldsmith-Pinkham

March 15, 2022

# Today's Topics

- Today, touching on two (related) topics
- First, finishing up conversation related to last class and diff-in-diff, focusing again on discussion of *event studies*
  - Focusing on how event studies generate a counterfactual control unit
- Second, discuss synthetic control (and dind) methods
  - Not completely new methods, but big upswing in research

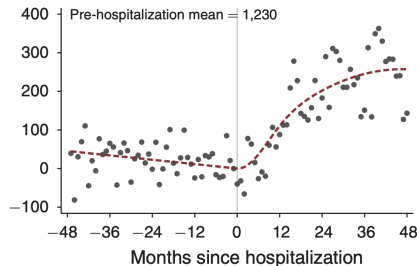
# Event study

- Distinction between “event study” and “diff-in-diff” is not formal
  - My view: event study are panel (or time series) settings where all units are eventually treated
- What does this imply? No group is a “pure control”
- Without a true control group, can't have both time fe, unit fe and the full non-parametric spec.

$$Y_{it} = \alpha_i + \gamma_t + \sum_{s=L_0, s \neq -1}^{L_1} 1(t - T_i = s) \mu_s$$

- Need to exclude both the baseline period AND at least some periods outside the treatment window

Panel B. Collection balances



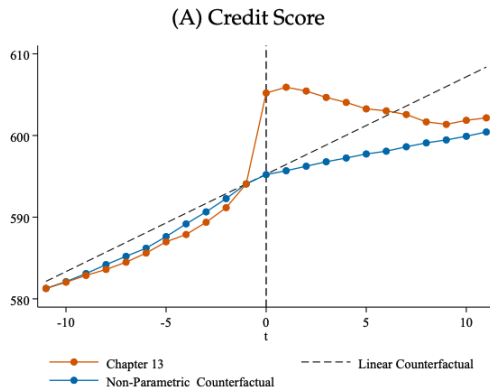
- Dobkin et al. (2018)
- Comparison is between those not yet hospitalized and those hospitalized

## Event study continued

- The necessary assumption is the same (or similar) what we discussed last class
  - Parallel trends – but amongst who?
  - Turns out, all of the groups need to be parallel.
  - That might be an awful assumption (e.g. very far apart from one another)
- Sun and Abraham (2020) point out that heterogeneity in treatment effects can have big impacts as well
  - In fact, it can violate the pre-trend test
  - The use of “excluded” periods potentially contaminates pre-periods
  - Solution: use “late” adopters as control group.
- Additional untestable assumptions are required as we allow for more types of heterogeneity

## Aside in event studies

- A key factor in how you construct your counterfactual (and what assumptions you find plausible) are a function of how far into the future you want to estimate outcomes
- An extremely short-run counterfactual could potentially just be a linear extrapolation
  - This assumes that the underlying model is locally linear, rather than globally
  - Construct a counterfactual from just a single time series, but highly non-robust
- Example from a robustness check in my own work (Dobbie et al. 2020)



# Constructing a counterfactual is the key goal

- Issue in event study was the attempt to get a “free lunch” – we always need a control group
- Think back to cross-sectional setting with ATT
  - We always knew  $Y_i(1)$ . Key issue is an estimator for  $Y_i(0)$ .
  - Event study approaches had issues by ignoring this point and hoping regression would solve problem
  - Notably, this problem disappears if we have full homogeneity + no anticipation and only exclude pre-periods
- Point of emphasis – we need parallel trends to hold to construct a counterfactual in these settings. Why?  $Y_{jt}(0) - Y_{j,t-1}(0)$  needs to be a good approximator of  $Y_{i,t}(0) - Y_{i,t-1}(0)$ .
  - Since we imposed  $Y_{it} = \alpha_i + \gamma_t + D_{it}\tau$ , the first differencing makes them good approximations

# Generalizing the Dind approach

- Pivoting slightly: instead of imposing the parallel trends assumption directly through the linear model, we could construct a combination of units to approximate  $Y_{it}(0)$ 
  - This is what one does in the cross-sectional setting with a pscore method! E.g. consider the ATT:

$$\tau_{ATT} = \underbrace{Y(1)}_{\text{Fully observed}} - \underbrace{\hat{Y}(0)}_{\text{Constructed}}$$

- How would one pick? Recall that with p-score methods or regression, weights effectively reweight based on comparability to treated group
  - With panel data, can use pre-treatment data to construct these weights
  - This method is known as synthetic control (and its various descendents)

# Synthetic Control example - (Abadie et al. (2010))

- Consider following problem: California bans smoking in 1989. What does that do to smoking?
  - Define estimand:  $\tau_{ban, CA} = Y_{california, post}(1) - Y_{california, post}(0)$
  - This is the effect of the *California* smoking ban
  - How can we get at it?
- We need a “synthetic California” as our control
- In an ideal world, the average of the other states would work – however, not clear empirically that they are a good counterfactual

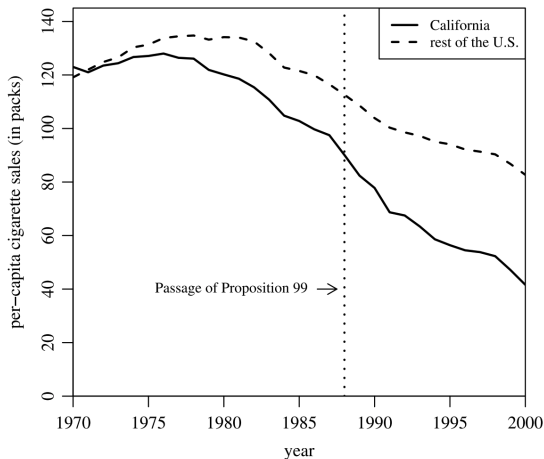


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.



## Generalized setup (Doudchenko and Imbens (2018))

- Consider the following general problem
- We have a panel with  $T$  time periods and  $N + 1$  units. Intervention  $D_{it}$  at time  $T_0$  for one unit (unit  $i = 0$ )
- Potential outcomes  $Y_{it}(D_{it})$ , and we only observe one of the potential outcomes (as per usual)
  - Fundamental problem of causal inference
  - We can also have fixed characteristics  $X_{it}$
- Let  $\mathbf{Y}_{a,b}$  denote the vector (or matrix in control case) for  $a \in \{\text{treatment, control}\}$  and  $b \in \{\text{pre, post}\}$  for the treated and control groups in the pre or post period.
- Then, we have observations (analogous setup for the covariates):

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_{t,post} & \mathbf{Y}_{c,post} \\ \mathbf{Y}_{t,pre} & \mathbf{Y}_{c,pre} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{t,post}(1) & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix}$$

## Generalized panel setup

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_{t,post} & \mathbf{Y}_{c,post} \\ \mathbf{Y}_{t,pre} & \mathbf{Y}_{c,pre} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{t,post}(1) & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix}$$

- To estimate  $\tau_i = Y_{t,post}(1) - Y_{t,post}(0)$ , we need an estimate for  $Y_{t,post}(0)$
- What if we just had the cross-section?
  - Note that if  $D_{it}$  were randomly assigned, we can derive an estimate using our p-score or regression methods
  - Even without random assignment, one could use covariates to match
  - Our main concern with p-score matching is bias
- Diff-in-diff exploited the panel structure by asserting a particular functional form

$$Y_{it} = \alpha_i + \gamma_t + D_{it}\tau + \epsilon_{it}$$

- Is there something particularly special about this linear additive factor structure?

## Generalized panel setup

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_{t,post} & \mathbf{Y}_{c,post} \\ \mathbf{Y}_{t,pre} & \mathbf{Y}_{c,pre} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{t,post}(1) & \mathbf{Y}_{c,post}(0) \\ \mathbf{Y}_{t,pre}(0) & \mathbf{Y}_{c,pre}(0) \end{pmatrix}$$

- Recall that our problem boils down to the estimate of an untreated “synthetic” unit
- Following Doudchenko and Imbens (2018), note estimators of the following form:

$$\hat{Y}_{t,post}(0) = \mu + \sum_{i \in \mathcal{C}} \omega_i Y_{i,T}$$

- A constant  $\mu$  allows for very different averages (common in diff-in-diff)
  - Weights are allowed to vary across  $i$  – a simple average would be diff-in-diff
- We can now consider deviations from diff-in-diff

# The synthetic control method (Abadie et al. (2010))

$$\hat{Y}_{t,post}(0) = \mu + \sum_{i \in \mathcal{C}} \omega_i Y_{i,T}$$

- In ADH, they impose
  1.  $\mu = 0$
  2.  $\sum_i \omega_i = 1$
  3.  $\omega_i \geq 0 \forall i$
- These three restrictions create a counterfactual California whose outcomes are within the support of the other states, and is a weighted sum of a subset of states

# The synthetic control method (Abadie et al. (2010))

$$\hat{Y}_{t,post}(0) = \mu + \sum_{i \in \mathcal{C}} \omega_i Y_{i,T}$$

- Formally, the  $\omega_i$  need to be estimated, and are constructed by minimizing the distance between covariates in the pre-period:

$$||\mathbf{X}_{\text{treat}} - \mathbf{X}_{\text{control}} \mathbf{W}||$$

- The crucial piece tying this together:  $\mathbf{X}$  can include both lagged outcomes, and covariates.
- Note we can now re-envision our panel data:
  - Observed outcomes:  $\mathbf{Y}_{t,post}(1), \mathbf{Y}_{c,post}(0)$
  - Observed covariates / predictors:  $\mathbf{Y}_{t,pre}(0), \mathbf{Y}_{c,pre}(0), \mathbf{X}_t, \mathbf{X}_c$
- In many ways, this is just a matching problem using many characteristics!

# The synthetic control method (Abadie et al. (2010))

$$\hat{Y}_{t,post}(0) = \mu + \sum_{i \in \mathcal{C}} \omega_i Y_{i,T}$$

- Formally, the  $\omega_i$  need to be estimated, and are constructed by minimizing the distance between covariates in the pre-period:

$$\{\hat{\omega}\}_i = \arg \min_{\mathbf{W}} ||\mathbf{X}_{treat} - \mathbf{X}_{control} \mathbf{W}||$$

- The crucial piece tying this together:  $\mathbf{X}$  can include both lagged outcomes, and covariates.
- Note we can now re-envision our panel data:
  - Observed outcomes:  $\mathbf{Y}_{t,post}(1), \mathbf{Y}_{c,post}(0)$
  - Observed covariates / predictors:  $\mathbf{Y}_{t,pre}(0), \mathbf{Y}_{c,pre}(0), \mathbf{X}_t, \mathbf{X}_c$
- In many ways, this is just a matching problem using many characteristics!

# The synthetic control method (Abadie et al. (2010))

- This approach can be incredibly successful

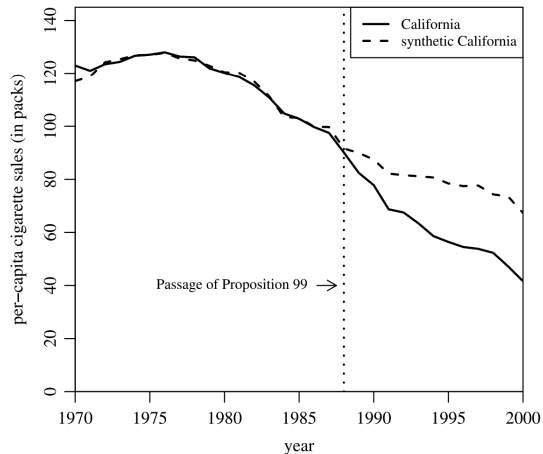


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

# The synthetic control method (Abadie et al. (2010))

- This approach can be incredibly successful
- By careful construction of a synthetic control, can calculate counterfactual impacts due to policy
- Still subject to same caveats from DiD
  - not invariant to some transformations (e.g. log and linear)

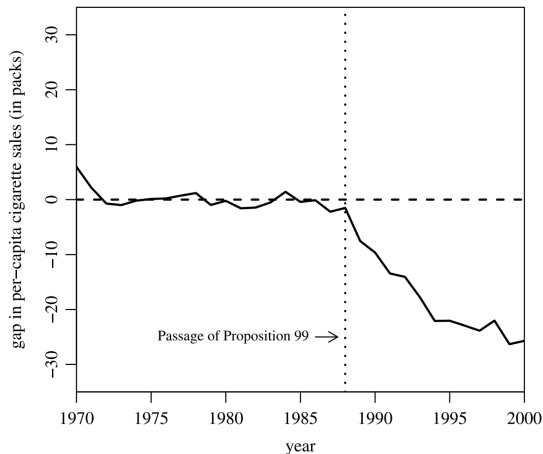


Figure 3. Per-capita cigarette sales gap between California and synthetic California.



# Inference in the synthetic control method (Abadie et al. (2010))

- Inference for this method is slightly more complex, as there is only a single treated unit
  - Large sample asymptotics unlikely to work
- Placebo approach is standard: apply method to each potential control unit, and report effect in period
- Analogy here is to a randomization inference argument, comparing to a “null” effect

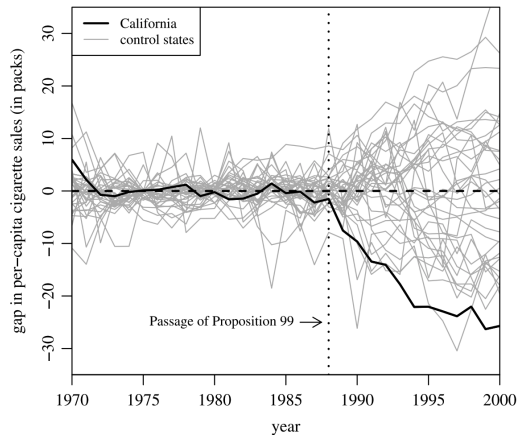


Figure 5. Per-capita cigarette sales gaps in California and placebo gaps in 34 control states (discards states with pre-Proposition 99 MSPE twenty times higher than California's).

## Synthetic Diff-in-diff

- In Arkhangelsky et al. (2019), they show you can rewrite the synthetic control estimator as

$$(\hat{\mu}, \hat{\gamma}, \hat{\tau}) = \arg \min_{\mu, \gamma, \tau} \sum_i \sum_t (Y_{it} - \mu - \gamma_t - D_{it}\tau)^2 \hat{\omega}_i,$$

subject to the  $\hat{\omega}_i$  chosen via the SC approach

- Contrast that with DID:

$$(\hat{\mu}, \hat{\alpha}, \hat{\gamma}, \hat{\tau}) = \arg \min_{\mu, \gamma, \tau} \sum_i \sum_t (Y_{it} - \mu - \alpha_i - \gamma_t - D_{it}\tau)^2$$

- They then propose a more robust approach, called Synthetic diff-in-diff, which estimates

$$(\hat{\mu}, \hat{\alpha}, \hat{\gamma}, \hat{\tau}) = \arg \min_{\mu, \gamma, \tau} \sum_i \sum_t (Y_{it} - \mu - \alpha_i - \gamma_t - D_{it}\tau)^2 \hat{\omega}_i \hat{\lambda}_t$$

- This approach relaxes the parallel trends assumption by requiring parallel trends in an underlying approximate factor structure

# Synthetic Diff-in-diff

- Key difference is twofold:
  1. Pre-trend means do not need to match “exactly”
  2. Weighting is not equivalent across all time periods
- Conceptually – different ways to generate the counterfactual given a model

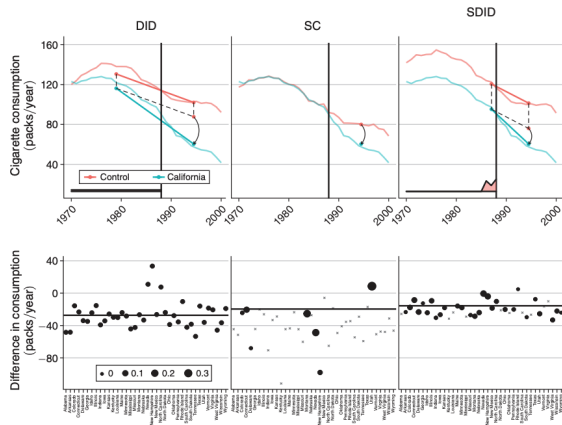


FIGURE 1. A COMPARISON BETWEEN DID, SC, AND SDID ESTIMATES FOR THE EFFECT OF CALIFORNIA PROPOSITION 99 ON PER-CAPITA ANNUAL CIGARETTE CONSUMPTION (IN PACKS/YEAR)

# Synthetic Diff-in-diff

- Key difference is twofold:
  1. Pre-trend means do not need to match “exactly”
  2. Weighting is not equivalent across all time periods
- Conceptually – different ways to generate the counterfactual given a model

$$\hat{\tau} = \hat{\delta}_{tr} - \sum_{i=1}^{N_{co}} \hat{\omega}_i \hat{\delta}_i \quad \text{where} \quad \hat{\delta}_{tr} = \frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N \hat{\delta}_i.$$

$$\hat{\delta}_i^{sc} = \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^T Y_{it},$$

$$\hat{\delta}_i^{did} = \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^T Y_{it} - \frac{1}{T_{pre}} \sum_{t=1}^{T_{pre}} Y_{it},$$

$$\hat{\delta}_i^{sdid} = \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^T Y_{it} - \sum_{t=1}^{T_{pre}} \hat{\lambda}_t^{sdid} Y_{it}.$$

# Synthetic Diff-in-diff

- So far, synth dind method discussion focused on single adoption period.
  1. Staggered adoption in synthetic control isn't meaningful
  2. How can you adopt it?
- Conceptually – split up the adoption timings a la Calloway & Sant'anna and others

$$\mathbf{W} = \begin{pmatrix} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 4 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 5 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 6 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

# Synthetic Diff-in-diff

- So far, synth dind method discussion focused on single adoption period.
  1. Staggered adoption in synthetic control isn't meaningful
  2. How can you adopt it?
- Conceptually – split up the adoption timings a la Calloway & Sant'anna and others

$$\mathbf{W}^1 = \begin{pmatrix} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 6 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$

$$\mathbf{W}^2 = \begin{pmatrix} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 4 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

## So what about synthetic methods?

- Many authors argue that this is a burgeoning innovative field
  - It is a very cool method!
- But, I am not sure I see it breaking a huge amount of ground
- Researchers appear hesitant to use it. Why?
- My thoughts:
  - These are strong structural assumptions, and not clear we have good tests yet
  - Despite concerns re: pre-trends in dind, the assumptions felt testable
- Researcher degrees of freedom seem multifold. True in DinD too, but perhaps more transparent?
  - More worrisome: dind is equally problematic, but we aren't aware of it

## My recommendation / takeaway

- Synthetic control is the ideal approach when faced with a single treatment
  - By far the most natural approach in this setting, and is a practical approach
  - Typical approach – get a good synthetic control for a given treatment. If none exists, stop. Ben-Michael, Feller and Rothstein (2021) provide a better approach, which adjusts for imperfect pre-match.
- Synthetic DiD seems very promising as a generalization
  - Key question is convincing readers why this should work better than traditional method
  - My view: empirical papers will first need to show how / why their method works with both diff-in-diff and synth diff-in-diff
- Key point: *all of this relies on a model of the control outcome*
- Three packages to explore: augsynth/tidysynth and synthdid packages (original synth package is tough to use)