Part C: Panel Data Methods

C4: Synthetic Control Methods and Factor Models

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2023

C4 outline

1. Synthetic control methods

2. Factor models

Recommended reading: Abadie (2021 JEL)

Setting

Consider a complete panel with a non-staggered binary treatment again

	$t = 1$	$\dots$	$t = T_0$	$t = T_0 + 1$	$\dots$	$t = T_0 + T_1 \equiv T$
$i = 1$						
$\dots$						
$i = N_0$						
$i = N_0 + 1$						
$\dots$						
$i = N_0 + N_1 \equiv N$						

Setting (2)

For now, assume $N_1 = T_1 = 1$:

	$t=1$	$\dots$	$t=T_0$	$t=T$
$i=1$				
$\dots$				
$i=N_0$				
$i=N$				

- We just need to impute $Y_{NT}(0)$ to get $\hat{\tau}_{NT} = Y_{NT} - \widehat{Y_{NT}(0)}$
- We'll come back to inference later

DiD vs. synthetic control method (SCM)

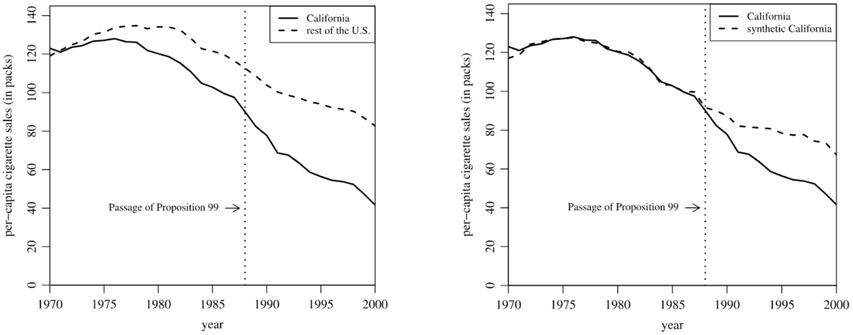
DiD uses the simple average of untreated units ("donors") as the control:

$$\hat{\tau}_{NT} = Y_{NT} - \frac{1}{N_0}\sum_{i=1}^{N_0} Y_{iT} - \frac{1}{T_0}\sum_{t=1}^{T_0}\left(Y_{Nt} - \frac{1}{N_0}\sum_{i=1}^{N_0} Y_{it}\right) \equiv Y_{NT} - \widehat{Y_{NT}(0)}$$

- What if we found a weighted average of donors that closely traced the pre-treatment path of Y_{Nt}: a "synthetic control" unit?
  - If the same relationship continues into $t = T$, can use $\widehat{Y_{NT}(0)} = \sum_{i=1}^{N_0} \omega_{iT} Y_{iT}$

Example: Abadie, Diamond, Hainmueller (2010)

Abadie et al. (2010) study the effect of California's 1988 tobacco control program (Proposition 99) on cigarette sales per capita



- PTA clearly fails
- Avoid manually picking comparable states (as in Card and Krueger 1994)

How does it work?

For some $\{v_t\}$, we choose weights ω_i on donors to solve

$$\min_{\omega_1,\ldots,\omega_{N_0}} \sum_{t=1}^{T_0} v_t \cdot \left(Y_{Nt} - \sum_{i=1}^{N_0} \omega_i Y_{it} \right)^2 \qquad \text{s.t. } \omega_i \geq 0, \quad \sum_{i=1}^{N_0} \omega_i = 1$$

- "Simplex constraints" produce a well-defined average and avoid extrapolation
- They are also a form of regularization
  - Otherwise, a **"vertical" regression** of Y_{Nt} on $Y_{1t}, \ldots, Y_{N_0 t}$ across $t = 1, \ldots, T_0$
    - ★ With $N_0 > T_0$ there would be ∞ ways to fit Y_{Nt} in pre-periods perfectly
  - And no reason to get good $\widehat{Y_{NT}(0)}$ — overfitting
  - With the constraints, *typically* get a unique, sparse solution: few non-zero weights
- Sparsity makes the counterfactual transparent

Details

1. Besides pre-period outcomes, can match on any predetermined predictors: $X_i = (Y_{i1}, \ldots, Y_{iT_0}, Z_i)$
2. How to pick weights on predictors, v?
   - Inverse variance of the predictor across all units
   - Or cross-validation:
     - Choose training period $t = 1, \ldots, t_0$
     - Search for v to minimize out-of-sample MSE (on the validation period $t_0 + 1, \ldots, T_0$)
     - For estimation, limit the sample to the last t_0 pre-periods & treated period

Details (2)

3. $\{\omega_i\}$ are not unique if the treated unit is in the convex hull of many donors

   - Abadie and L'Hour (2019): try to match with donors that are more similar to the treated unit

     $$\min_{\omega_1,\ldots,\omega_{N_0}} \sum_t \left\{ \left(Y_{Nt} - \sum_{i=1}^{N_0} \omega_i Y_{it} \right)^2 + \lambda \sum_{i=1}^{N_0} \omega_i \left(Y_{Nt} - Y_{it} \right)^2 \right\}$$

     s.t. $\omega_i \geq 0$, $\sum_{i=1}^{N_0} \omega_i = 1$, with penalty $\lambda > 0$

     - ⋆ Restores uniqueness and reduces interpolation bias

   - Robbins, Saunders, Kilmer (2017): pick weights that minimize the distance from equal weights

Abadie et al. 2010 example (cont.)

As predictors, use several observables (averaged in 1980–88) and outcome in three pre-periods

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:* Drop states that adopted other tobacco restrictions during the sample period

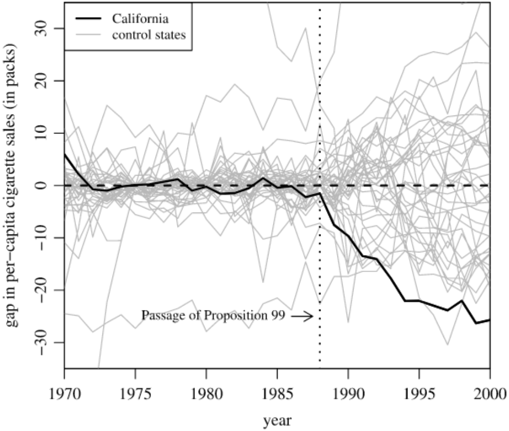
Synthetic California (Abadie et al. 2010)

State	Weight
Utah	0.334
Nevada	0.234
Montana	0.199
Colorado	0.164
Connecticut	0.069
Other 33 states	0

Inference

- Inference is difficult with only 1 treated unit
- Abadie, Diamond, Hainmueller (2010): randomization inference (RI)
    - Under the null of zero effect, the treated unit is no different than others
    - For each i (including $i = N$), compute $R_i = \sum_{t>T_0} \left(Y_{it} - \widehat{Y_{it}(0)} \right)^2$
    - Reject if R_N is extreme in the set of $R_1, \ldots, R_N$; p-value $= \frac{1}{N} \sum_{i=1}^{N} \mathbf{1} \left[R_N \geq R_i \right]$

Inference in Abadie et al. (2010)



Inference details

- Complication 1: for some i, synthetic control may not match the pre-treatment trajectory well
  - Can drop units with high pre-treatment MSE
  - Or use $R_i =$ post-treatment MSE divided by pre-treatment MSE
- Complication 2: how can we get a confidence interval?
  - Test inversion: for each τ_0 test $\tau = \tau_0$; collect all τ_0 that are not rejected
  - To test $\tau = \tau_0$, replace Y_{NT} with $Y_{NT} - \tau_0$ and test $\tau = 0$

Inference details (2)

- Problem: RI is not justified without randomization
  - Test has a 5% significance level in the sense that for 5% of units if they were treated the correct null would be rejected
    - Not very useful since the treated unit was not chosen randomly with equal probabilities
  - Abadie, Diamond, Hainmueller (2015) study the economic impacts of German reunification on West Germany
    - What does assigning this treatment to another country mean?
- One alternative (Chernozhukov, Wuthrich, Zhu 2021): "conformal inference" based on permuting residuals for the treated unit across *periods*

Diagnostic testing and robustness

- Placebo test: **"backdating"**

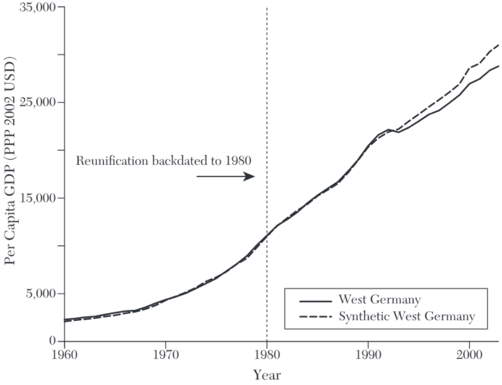


Figure 3. Backdating the 1990 German Reunification Application

(From Abadie 2021)

- Robustness check: dropping each contributing donor at a time

Some extensions

- Everything works with multiple treated periods
- Multiple treated units:
  - Create a synthetic control for each treated unit and estimate the effects, then average (Abadie and L'Hour 2019)
  - Or a single synthetic control for the average of treated units (Robbins, Saunders, Kilmer 2017)
- Sun, Ben-Michael, Feller (October 10, 2023): construct a single synthetic control for multiple outcome variables
- Ben-Michael, Feller, Rothstein (2021): Augmented SCM estimators *(augsynth)*
  - Adjust the estimator for mismatches in the pre-periods for double-robustness

SCM: When and why?

In which contexts should one use synthetic controls?

- Abadie (2021):
  - No anticipation, no spillovers
  - Donor units shouldn't experience large idiosyncratic shocks in the post-periods
  - Big effects and low-volatility outcomes (when few treated units)
  - When a good synthetic control exists
- But which outcome models imply existence of good synthetic controls?
  - And is SCM the best for those models?
  - SCM is usually motivated by factor models...

Outline

Factor model

Factor (a.k.a. latent factor, interactive fixed-effect) model:

$$Y_{it}(0) = \alpha_i + \beta_t + L_{it} + \varepsilon_{it}, \qquad L_{it} = \sum_{r=1}^{R} S_{ri} F_{rt}, \qquad \mathbb{E}\left[\varepsilon_{it}\right] = 0$$

with an unknown small number R of unobserved time-varying factors F_{rt} and unobserved unit sensitivities (factor loadings) S_{ri}

- $\alpha_i = \alpha_i \cdot 1$; $\beta_t = 1 \cdot \beta_t$; unit-specific trend $= \gamma_i \cdot t$
- Viewing S_{ri} and F_{rt} as parameters, this is a nonlinear model
- Treatment D_{it} can be correlated with L_{it} (but not with ε_{it})

Factor models vs. SCM

$$Y_{it}(0) = \alpha_i + \beta_t + L_{it} + \varepsilon_{it}, \qquad L_{it} = \sum_{r=1}^{R} S_{ri} F_{rt}, \qquad \mathbb{E}\left[\varepsilon_{it}\right] = 0$$

Why synthetic controls if you have a factor model?

- $Y_{Nt} \approx \sum_{i=1}^{N_0} \omega_i Y_{it}$ for many pre-periods t if $S_{rN} \approx \sum_{i=1}^{N_0} \omega_i S_{ri}$ for each factor r
- Then can recover $\hat{\alpha}_i + \hat{\beta}_t + \hat{L}_{NT} \equiv \widehat{Y_{NT}(0)}$ with SCM... but there are better methods

Strategy #1: Synthetic DiD (Arkhangelsky et al. 2021)

- A modern take on synthetic controls and DiD. Balances out latent factors
- Diff-in-diff puts equal weight on all untreated units and all pre-periods
- But control units similar to the treated ones in the pre-periods are more useful
  - Must have factor loadings similar to $i = N$ (or to average of $N_0 + 1, \ldots, N$)
- Pre-periods similar to the post-period on untreated outcomes are more useful
  - Must have factors similar to $t = T$ (or to average of $T_0 + 1, \ldots, T$)
- Obtain weights $v_1, \ldots, v_{T_0}$ and $\omega_1, \ldots \omega_{N_0}$ that add up to one; then

$$\widehat{ATT} = \overline{Y}_{\text{treated,post}} - \sum_{i=1}^{N_0} \hat{\omega}_i \overline{Y}_{i,\text{post}} - \sum_{t=1}^{T_0} \hat{v}_t \left(\overline{Y}_{\text{treated},t} - \sum_{i=1}^{N_0} \hat{\omega}_i Y_{it} \right)$$

- Implemented as TWFE regression weighted by $\hat{v}_t \cdot \hat{\omega}_i$ (with $\hat{v}_{T_0+1} = \ldots = \hat{v}_T = 1$ and $\hat{\omega}_{N_0+1} = \ldots = \hat{\omega}_N = 1$)

How to choose $\hat{v}_t$ and $\hat{\omega}_i$?

- Choose $\hat{v}_t$ to make $\overline{Y}_{i,\text{post}}$ close to $c_1 + \sum_{t=1}^{T_0} v_t Y_{it}$ across untreated units
  - c_1 captures period FEs
  - Vertical regression of $\overline{Y}_{i,\text{post}}$ on a constant and $Y_{i1}, \ldots, Y_{iT_0}$ with restrictions $\sum_{t=1}^{T_0} v_t = 1$ and $v_t \geq 0$
- Choose $\hat{\omega}_i$ to make $\overline{Y}_{\text{treated},t}$ close to $c_2 + \sum_{i=1}^{N_0} \omega_i Y_{it}$ across periods
  - c_2 captures unit FEs
  - Horizontal regression; because $N_0 > T_0$, regularize (using ridge)

$$\min_{c_2,\{\omega_i\}} \frac{1}{T_0} \sum_{t=1}^{T_0} \left(\overline{Y}_{\text{treated},t} - c_2 - \sum_{i=1}^{N_0} \omega_i Y_{it} \right)^2 + \lambda \sum_{i=1}^{N_0} \omega_i^2 \qquad \text{s.t. } \sum_{i=1}^{N_0} \omega_i = 1, \quad \omega_i \geq 0$$

(See paper for choice of penalty λ)

Inference in synthetic diff-in-diff

Arkhangelsky et al. (2021) propose:

- With sufficiently many treated units:
  - Unit-clustered bootstrap
  - Jackknife approach based on how $\widehat{ATT}$ varies when dropping units one at a time (without reestimating weights)
- With few treated units:
  - Placebo approach: variation across placebo estimates that assign treatment to random N_1 of untreated units

Strategy #2: Imputation with factor models

- Recall $Y_{it}(0) = \alpha_i + \beta_t + L_{it} + \varepsilon_{it}, \quad L_{it} = \sum_{r=1}^{R} S_{ri} F_{rt}$
- SCM-type methods balance out latent factors without estimating them
- Alternatively, can estimate L_{it} from untreated observations and use them for $\widehat{Y_{it}(0)}$
- Can estimate factors, loadings, and R using principle component analysis (Bai and Ng 2002)
- Better idea: estimate the L matrix directly
    - Key property: $\text{rank}(L) = R$ is small

Athey et al. (2021) matrix completion

Athey, Bayati, Doudchenko, Imbens, Khosravi (2021): **matrix completion** approach

L_{it}	$t=1$	$\dots$	$t=T_0$	$t=T_0+1$	$\dots$	$t=T_0+T_1\equiv T$
$i=1$						
$\dots$						
$i=N_0$						
$i=N_0+1$				?	?	?
$\dots$				?	?	?
$i=N_0+N_1\equiv N$				?	?	?

- Recover L from:
  - observing Y_{it} for untreated observations = noisy version of $\alpha_i+\beta_t+L_{it}$
  - knowing that L has low rank

Athey et al. (2021) matrix completion

Athey et al. (2021) solve:

$$\min_{\{\alpha_i\},\{\beta_t\},L} \sum_{it:\ D_{it}=0} \left(Y_{it} - \alpha_i - \beta_t - L_{it}\right)^2 + \lambda \left\|L\right\|_*$$

- $\left\|L\right\|_*$ is the "nuclear" norm of matrix L: small when L is close to low-rank
- Computationally efficient (unlike searching among low-rank matrices)
- Then take average of $Y_{it} - \hat{\alpha}_i - \hat{\beta}_t - \hat{L}_{it}$ among treated observations
- No results on inference