Synthetic Control Method

Kosuke Imai

Harvard University

Spring 2021

Synthetic Control Method (Abadie et al. 2010. *J. Am. Stat. Assoc.*)

- Setting: N units and T time periods
- One treated unit $i = N$ receiving the treatment at time T
- Quantity of interest: $Y_{NT}(1) - Y_{NT}(0) = Y_{NT} - Y_{NT}(0)$
- Create a synthetic control using past outcomes
- Weighted average:

$$\widehat{Y_{NT}(0)} = \sum_{i=1}^{N-1} \hat{w}_i Y_{iT}$$

where the weights balance past outcomes

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{t=1}^{T-1} \left(Y_{Nt} - \sum_{i=1}^{N-1} w_i Y_{it} \right)^2$$

with $\sum_{i=1}^{N-1} \hat{w}_i = 1$ and $\hat{w}_i \geq 0$

Causal Effect of ETA's Terrorism

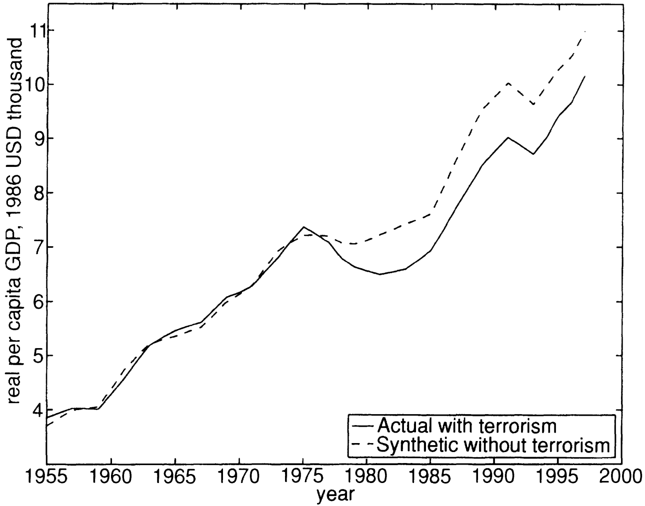


FIGURE 1. PER CAPITA GDP FOR THE BASQUE COUNTRY

(Abadie and Gardeazabal. 2003. *Am. Econ. Rev.)*

Model-based Justification

- The main motivating factor analytic model:

$$Y_{it}(0) \quad = \quad \gamma_t + \delta_t^\top \mathbf{X}_i + \xi_t^\top \mathbf{U}_i + \epsilon_{it}$$

  - Generalization of the linear two-way fixed effects model
  - Key assumption: there exist weights such that

$$\sum_{i=1}^{N-1} w_i \mathbf{X}_i = \mathbf{X}_N \quad \text{and} \quad \sum_{i=1}^{N-1} w_i \mathbf{U}_i = \mathbf{U}_N$$

- Another motivating autoregressive model with time-varying covariates:

$$\begin{aligned} Y_{it}(0) &= \rho_t Y_{i,t-1}(0) + \delta_t^\top \mathbf{X}_{it} + \epsilon_{it} \\ \mathbf{X}_{it} &= \lambda_{t-1} Y_{i,t-1}(0) + \Delta_{t-1} \mathbf{X}_{i,t-1} + \nu_{it} \end{aligned}$$

  - Past outcomes can affect current treatment
  - No unobserved time-invariant confounders

Synthetic Control with Pre-treatment Covariates

- We can generalize the synthetic control method
- Pre-treatment covariates: $\mathbf{Z}_i = (\mathbf{Y}_i^\top, \mathbf{X}_i^\top)^\top$
  - lagged outcomes: $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i,T-1})^\top$
  - lagged covariates $\mathbf{X}_i = (\mathbf{X}_{i1}^\top, \mathbf{X}_{i2}^\top, \ldots, \mathbf{X}_{i,T-1}^\top)^\top$
- Or some subsets or functions of these variables

- Balance both the lagged outcomes and pre-treatment covariates

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \left(\mathbf{Z}_N - \sum_{i=1}^{N-1} w_i \mathbf{Z}_i\right)^\top \widehat{\Sigma}^{-1} \left(\mathbf{Z}_N - \sum_{i=1}^{N-1} w_i \mathbf{Z}_i\right)$$

$$\text{subject to } \sum_{i=1}^{N-1} w_i = 1, \text{ and } w_i \geq 0 \text{ for all } i = 1, \ldots, N-1,$$

where $\widehat{\Sigma}$ is the covariance matrix of $\mathbf{Z}_i$

Placebo Test

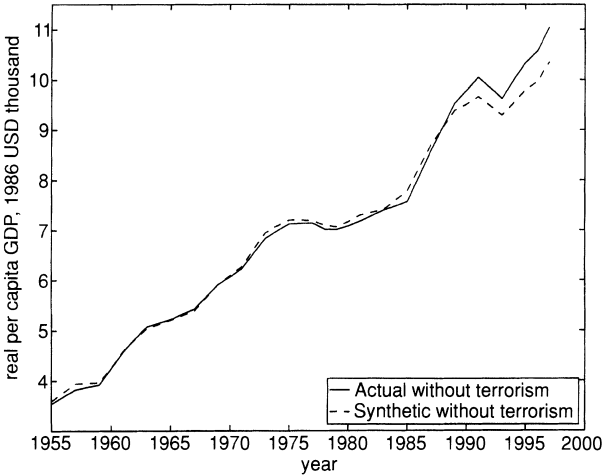
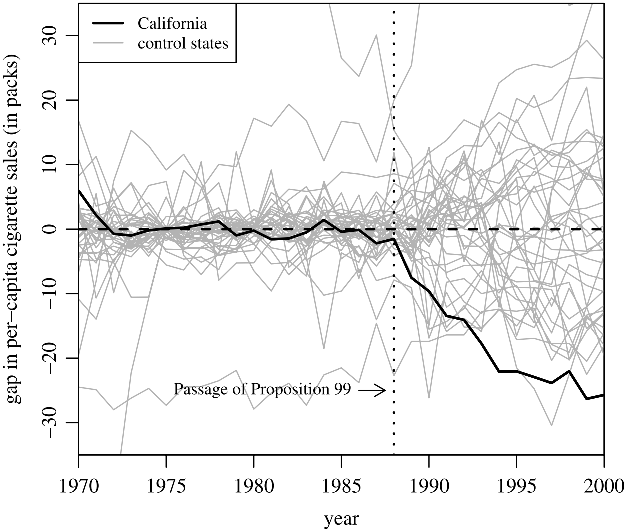


FIGURE 4. A "PLACEBO STUDY," PER CAPITA GDP FOR CATALONIA

can do this for all control units and compare them with the treated unit

Permutation Test

- Assumption: Errors are exchangeable



- We can invert this test to obtain a point-wise confidence interval

Relationship with Regression

- The synthetic control can be seen as a constrained regression:

$$(\hat{\alpha}, \mathbf{w}) = \underset{\alpha, \mathbf{w}}{\operatorname{argmin}} \sum_{t=1}^{T-1} \left(Y_{Nt} - \alpha - \sum_{i=1}^{N-1} w_i Y_{it} \right)^2$$

  subject to the following constraints
  1. zero intercept: $\alpha = 0$
  2. positive weights $w_i \geq 0$ for all $i = 1, \ldots, N-1$
  3. sum-to-one constraint: $\sum_{i=1}^{N-1} w_i = 1$

- No time-invariant difference between the treated and control units
- Treated unit in the convex hull of the control units
- Regularization required when N is large relative to T