Synthetic Controls 合成控制

Lingguo Cheng

Nanjing University

December 22, 2025

Outline

1. Synthetic Controls

Synthetic Controls: Introduction

- The synthetic control method was originally developed for case study setups, with the goal to assess the treatment effect on a single treated unit based on a comparison with multiple nontreated units.
- This appears particularly well suited for the evaluation of relatively rare (or even unique) policy interventions that affect only one specific region, country, organization, or company.
- For instance, Abadie, Diamond, and Hainmueller (2015) apply the synthetic control method to evaluate the effect of the reunification of West and East Germany in 1990 on the development of the gross domestic product (GDP) per capita in West Germany.

- The idea of the synthetic control method is to impute the counterfactual outcome under nontreatment based on an appropriate combination of other countries that are sufficiently similar in economic terms to West Germany prior to the reunification, but did not experience any reunification or similar kind of policy change.

- More concisely, the treated unit's potential outcome under nontreatment (i.e., the West German counterfactual after the reunification) is synthetically estimated as a weighted average of the observed posttreatment outcomes coming from a donor pool of nontreated units, in this case other OECD countries.

- The weight that a nontreated country receives for computing this average called the synthetic control depends on how similar the country was to West Germany in terms of economic conditions in the pretreatment periods: that is, prior to reunification.

Example: Abadie and Gardeazabal (2003)

- Abadie and Gardeazabal (2003) consider the effect of the terrorist conflict in the Basque Country, a region in Spain, on GDP per capita in that region based on a synthetic control created from other Spanish regions.
- The start of the treatment is marked by 1968, when the terrorist group ETA claimed its first victim.
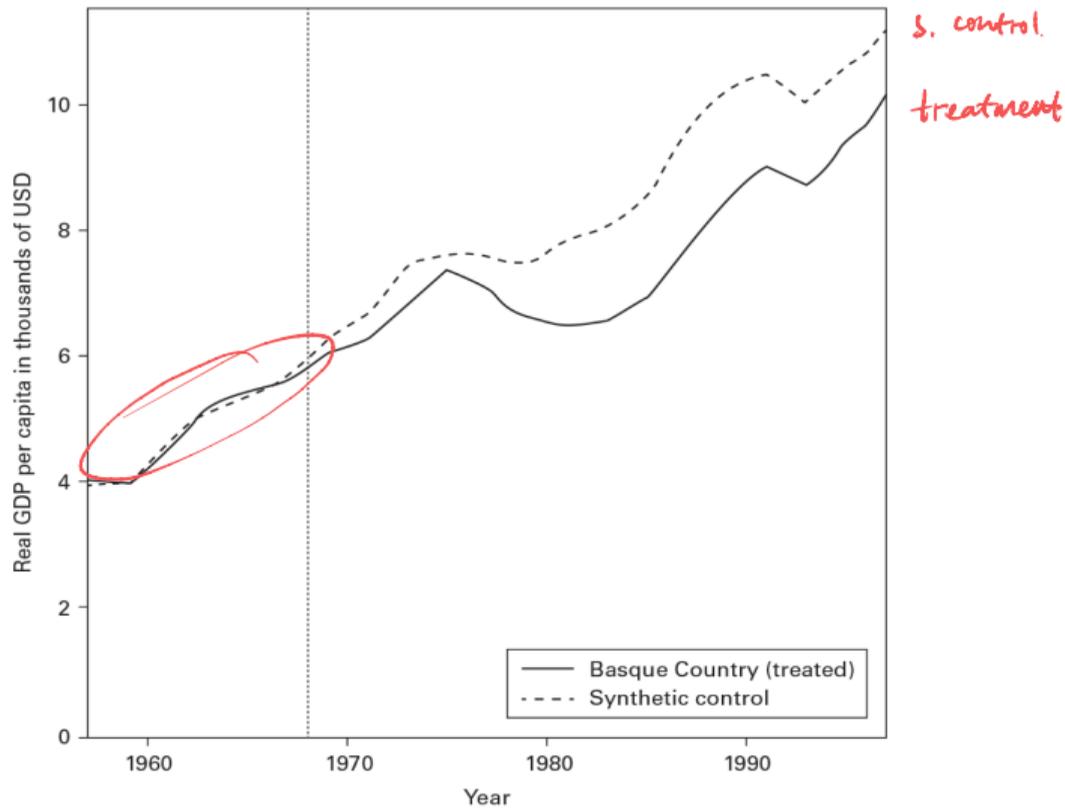- After the mid-1970s, when ETA's terrorist activities intensified

Figure 1: Synthetic control method for the terrorist conflict in Basque Country.

Formal Estimation and Inference

- A panel data set of n units, such that a unit's index $i \in \{1, \ldots, n\}$, time index $t \in \{1, \ldots, T\}$.
- Denote by Y_{it} the observed outcome of unit i in period t. Assume (without loss of generality) that only the last unit $i = n$ in the data is treated in a period $T_0 + 1$, with T_0 denoting the last period prior to treatment, which satisfies $T_0 \geq 1$. Therefore, the treatment takes place in a period after the first period, implying that $1 < T_0 + 1 \leq T$.
- For any post-treatment period $t \geq T_0 + 1$, an estimate of the treatment effect for the treated unit $i = n$, denoted by Δ_{n, T_0+1}, as the difference of the treated outcome and a weighted average of control outcomes in that period:

$$\Delta_{n,t} = Y_{nt} - \sum_{i=1}^{n-1} \hat{w}_i Y_{it} \tag{1}$$

估计一套权重

where $\hat{w}_i$ is a specific weight or importance of some nontreated unit.

How to choose the weights: Version 1

- In its most basic version, the idea of the synthetic control method is to choose the weights such that the weighted average of pretreatment outcomes of nontreated units matches the evolution of the pretreatment outcome of the treated observation: that is, up to period T_0.
- This approach assumes that we can appropriately model the treated unit's posttreatment potential outcome under nontreatment by means of the reweighted outcomes of actually nontreated units.
- Put differently, controlling for pretreatment outcomes Y_{it} over periods $t \in \{1, \ldots, T_0\}$ must be sufficient to control for confounders entailing diverging potential outcomes under nontreatment of the treated and nontreated units in the posttreatment periods.
- Formally, we choose the weights such that

$$\sum_{i=1}^{n-1} \hat{\omega}_i Y_{it} \approx Y_{nt} \quad \text{for all } t = 1, \ldots, T_0 \tag{2}$$

Comments

- This approach seems quite related to a selection-on-observables framework, with pretreatment outcomes serving as covariates to be controlled for.
- Like the selection-on-observables framework, the synthetic control method relies on a type of common support condition concerning the existence of treated and nontreated units that are similar in terms of pretreatment outcomes, namely, the so-called ***convex hull condition***.
- The latter implies that the pretreatment outcomes of the treated unit are not too extreme compared to the nontreated units—in particular, not much higher than the highest or much lower than the lowest outcome of nontreated units in any pretreatment period.
- Furthermore, anticipation effects must be ruled out, such that the pretreatment outcomes of the treated unit are not affected by the treatment being introduced.

Choosing the weight: (prototypical approach) Version 2

- More formally, the collection of weights $\hat{\omega} = (\hat{\omega}_1, \ldots, \hat{\omega}_{n-1})$ is computed based on the following least squares approach, with ω^* denoting candidate values for the weights:

$$\hat{\omega} = \arg\min_{\omega^*} \sum_{t=1}^{T_0} \left(Y_{nt} - \sum_{i=1}^{n-1} \omega_i^* Y_{it} \right)^2 \tag{3}$$

subject to:

$$\sum_{i=1}^{n-1} \omega_i^* = 1$$
$$\omega_i^* \geq 0 \quad \text{for all } i = 1, \ldots, n-1$$

Choosing the weight: Version 3

- All of the treated unit's pretreatment outcomes are considered equally important for creating a synthetic control that permits assessing the causal effect.
- In practice, however, we might want to give a different importance to different periods, such as giving a higher weight to more recent pretreatment periods.
- Accounting for a differential importance of periods is easily obtained by adding a period-dependent weight, denoted by $v_t \geq 0$, that takes higher values for more important periods and a value of zero for periods that should not be considered at all:

$$\hat{\omega} = \arg\min_{\omega^*} \sum_{t=1}^{T_0} v_t \left(Y_{nt} - \omega_1^* Y_{1t} - \ldots - \omega_2^* Y_{n-1t} \right)^2 \tag{4}$$

subject to: $\omega_i^* \geq 0, \quad \sum_{i=1}^{n-1} \omega_i^* = 1$

Other Variants

- We may transform outcomes prior to running the synthetic control method if this appears appropriate.
    - The logarithm of an outcome
    - Subtract the pretreatment mean in outcomes of a unit i from unit i's outcome in some period t, in order to generate demeaned outcomes:

$$\tilde{Y}_{it} = Y_{it} - \frac{1}{T_0} \sum_{\tau=1}^{T_0} Y_{it} \qquad (5)$$

- We may not only include pretreatment outcomes, but also (or even exclusively) observed covariates X as control variables when determining the weights of nontreated units, to generate a synthetic control that is similar to the treated unit in terms of those covariates.

Inference

- Randomization inference based on permutation (see Abadie, Diamond, and Hainmueller (2010)), which is based on estimating placebo effects among nontreated units for whom the true effect is known to be zero –In the spirit of bootstrap

- An alternative permutation method for computing p-values is conformal inference; see Chernozhukov, Wüthrich, and Zhu (2021)

- We may also apply the synthetic control method or related estimates with multiple treated units to estimate their ATET in a specific outcome period, denoted by $\hat{\Delta}_{D=1,T=t} = E\left[Y_t(1) - Y_t(0) \mid D = 1\right]$

- One obvious approach is to apply the estimator separately to each of the treated units, and then average over the effects to estimate the ATET:

$$\hat{\Delta}_{D=1,T=t} = \frac{1}{\sum_{i=1}^{n} D_i} \sum_{i:D_i=1} \hat{\Delta}_{i,T=t} \qquad (6)$$

with $\hat{\Delta}_{i,T=t}$ being the effect estimate for a treated unit i based on the nontreated door pool.

An alternative approach for multiple treated units

- An alternative approach is to estimate the weights in a way that equation (3) holds on average for the treated units rather than for each treated unit separately.
- To discuss this possibility more formally, denote by $n_1 = \sum_{i=1}^{n} D_i$ the number of treated units and assume (without loss of generality) that the observations are ordered in a way that all $n - n_1$ nontreated units appear at the top (or come first) and all treated units at the bottom (or come last) in the data.

$$\sum_{i=1}^{n-n_1} \hat{\omega}_i Y_{it} + \hat{\alpha} \approx \frac{1}{n_1} \sum_{j=n-n_1+1}^{n} Y_{jt}, \quad \text{for all } t = 0, \ldots, T_0, \qquad (7)$$

where allowing the constant term $\hat{\alpha}$ to be nonzero corresponds to a DiD-rather than selection-on-observesble-type synthetic control method.

- The weights will be obtained by (Arkhangelsky et al., 2019)

$$(\hat{\omega}, \hat{\alpha}) = \arg\min_{\omega^*, \alpha^*} \sum_{t=1}^{T_0} \left(\frac{1}{n_1} \sum_{j=n-n_1+1}^{n} (Y_{jt} - \alpha^* - \omega_1^* Y_{1t} - \cdots - \omega_2^* Y_{n-n_1 t}) \right)^2 \tag{8}$$

subject to:

$$\omega_i^* \geq 0, \quad \sum_{i=1}^{n-1} \omega_i^* = 1$$

Several Comments

- This synthetic DiD approach, relies on the assumption that a weighted average of the nontreated outcomes in the pretreatment periods permits replicating the trend of the average potential outcome under nontreatment among the treated units.

- This condition about the average trend still permits the outcome trends to differ on the individual level of each unit. Constituting an alternative estimation strategy to those for the selection-on-observables and common trend assumptions, respectively.

- For the case of multiple treated units, we may combine synthetic control with other (e.g., regression) methods, with the weights of each method being determined by cross-validation.

- For a comprehensive discussion of the synthetic control method, see the review article by Abadie (2021).

Thank You for Your Attention!

Questions & Comments:

chenglingguo@nju.edu.cn