Synthetic Control

Luiz Renato Lima (The University of Tennessee)

Time Series Econometrics - Fall 2020

Motivation

- Athey and Imbens describe synthetic controls as the most important innovation in the policy evaluation literature in the last 15 years
- it has been applied to study:
  1. the effects of right-to-carry laws (Donohue et al., 2017)

Motivation

- Athey and Imbens describe synthetic controls as the most important innovation in the policy evaluation literature in the last 15 years
- it has been applied to study:
  1. the effects of right-to-carry laws (Donohue et al., 2017)
  2. legalized prostitution (Cunningham and Shah, 2018)

Motivation

- Athey and Imbens describe synthetic controls as the most important innovation in the policy evaluation literature in the last 15 years
- it has been applied to study:
  1. the effects of right-to-carry laws (Donohue et al., 2017)
  2. legalized prostitution (Cunningham and Shah, 2018)
  3. immigration policy (Bohn et al., 2014),

Motivation

- Athey and Imbens describe synthetic controls as the most important innovation in the policy evaluation literature in the last 15 years
- it has been applied to study:
  1. the effects of right-to-carry laws (Donohue et al., 2017)
  2. legalized prostitution (Cunningham and Shah, 2018)
  3. immigration policy (Bohn et al., 2014),
  4. organized crime (Pinotti, 2015)

Motivation

- Athey and Imbens describe synthetic controls as the most important innovation in the policy evaluation literature in the last 15 years
- it has been applied to study:
    1. the effects of right-to-carry laws (Donohue et al., 2017)
    2. legalized prostitution (Cunningham and Shah, 2018)
    3. immigration policy (Bohn et al., 2014),
    4. organized crime (Pinotti, 2015)
    5. minimum wages (Allegretto et al., 2017; Jardim et al., 2017)

Motivation

- Only one aggregate unit, such as a state, or a school district, is exposed to an event or intervention of interest.
- For example, Abadie et al. (2010) study the effect of a large tobacco-control program adopted in California in 1988;
- Bifulco et al. (2017) evaluate the effects of an educational program adopted in the Syracuse, NY, school district in 2008.

Model Setting

- Suppose that we obtain data for $J+1$ units: $j = 1, 2, ..., J+1$.

Model Setting

- Suppose that we obtain data for $J+1$ units: $j = 1, 2, ..., J+1$.
- we assume that the first unit ($j = 1$) is the treated unit, that is, the unit affected by the policy intervention of interest

Model Setting

- Suppose that we obtain data for $J+1$ units: $j = 1, 2, ..., J+1$.
- we assume that the first unit ($j = 1$) is the treated unit, that is, the unit affected by the policy intervention of interest
- The "donor pool", that is, the control group, $j = 2, ..., J+1$ is a collection of untreated units, not affected by the intervention

Model Setting

- Suppose that we obtain data for $J + 1$ units: $j = 1, 2, ..., J + 1$.
- we assume that the first unit ($j = 1$) is the treated unit, that is, the unit affected by the policy intervention of interest
- The "donor pool", that is, the control group, $j = 2, ..., J + 1$ is a collection of untreated units, not affected by the intervention
- We assume that our data span T periods and that the first T_0 periods are before the intervention

Model Setting

- Suppose that we obtain data for $J+1$ units: $j = 1, 2, ..., J+1$.
- we assume that the first unit ($j = 1$) is the treated unit, that is, the unit affected by the policy intervention of interest
- The "donor pool", that is, the control group, $j = 2, ..., J+1$ is a collection of untreated units, not affected by the intervention
- We assume that our data span T periods and that the firrst T_0 periods are before the intervention
- For each unit, j, and time, t, we observe the outcome of interest, Y_{jt}.

Model Setting

- Suppose that we obtain data for $J+1$ units: $j = 1, 2, ..., J+1$.
- we assume that the first unit ($j = 1$) is the treated unit, that is, the unit affected by the policy intervention of interest
- The "donor pool", that is, the control group, $j = 2, ..., J+1$ is a collection of untreated units, not affected by the intervention
- We assume that our data span T periods and that the firrst T_0 periods are before the intervention
- For each unit, j, and time, t, we observe the outcome of interest, Y_{jt}.
- For each unit, j, we also observe a set of k predictors of the outcome, $X_{1j}...X_{kj}$, which may include pre-intervention values of Y_{jt}

Model Setting

- Suppose that we obtain data for $J+1$ units: $j = 1, 2, ..., J+1$.
- we assume that the first unit ($j = 1$) is the treated unit, that is, the unit affected by the policy intervention of interest
- The "donor pool", that is, the control group, $j = 2, ..., J+1$ is a collection of untreated units, not affected by the intervention
- We assume that our data span T periods and that the firrst T_0 periods are before the intervention
- For each unit, j, and time, t, we observe the outcome of interest, Y_{jt}.
- For each unit, j, we also observe a set of k predictors of the outcome, $X_{1j} ... X_{kj}$, which may include pre-intervention values of Y_{jt}
- The $k \times 1$ vectors $X_1, ..., X_{J+1}$contain the values of the predictors for units $j = 1, ..., J+1$

Model Setting

- Suppose that we obtain data for $J+1$ units: $j = 1, 2, ..., J+1$.
- we assume that the first unit ($j = 1$) is the treated unit, that is, the unit affected by the policy intervention of interest
- The "donor pool", that is, the control group, $j = 2, ..., J+1$ is a collection of untreated units, not affected by the intervention
- We assume that our data span T periods and that the firrst T_0 periods are before the intervention
- For each unit, j, and time, t, we observe the outcome of interest, Y_{jt}.
- For each unit, j, we also observe a set of k predictors of the outcome, $X_{1j}...X_{kj}$, which may include pre-intervention values of Y_{jt}
- The $k \times 1$ vectors $X_1, ..., X_{J+1}$contain the values of the predictors for units $j = 1, ..., J+1$
- The $k \times J$ matrix $X_0 = [X_2 ... X_{J+1}]$ collects the values of the predictors for the J untreated units.

Model Setting

- For each unit, j, and time period, t, we defite Y_{jt}^N to be the potential response without intervention.
- For the unit affected by the intervention, $j = 1$, and a post-intervention period, $t > T_0$, we define Y_{1t}^I to be the potential response under the intervention.
- Then, the effect of the intervention of interest for the affected unit in period t (with $t > T_0$) is:

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

Challenge

$$\tau_{1t} = Y_{1t}^{I} - Y_{1t}^{N} \tag{1}$$

- Because unit "one" is exposed to the intervention after period T_0, it follows that for $t > T_0$ we have $Y_{1t} = Y_{1t}^{I}$
- The great policy evaluation challenge is to estimate the counterfactual Y_{1t}^{N} for $t > T_0$.
- Notice that (1) allows the effect of the intervention to change in time.
- intervention effects may not be instantaneous and may accumulate or dissipate as time after the intervention passes.

Estimation

- A synthetic control (Y_{1t}^N) is defined as a weighted average of the units in the control group

$$\begin{aligned}
\widehat{Y}_{1t}^N &= \sum_{j=2}^{J+1} \omega_j Y_{jt} \\
\tau_{1t} &= Y_{1t}^I - \widehat{Y}_{1t}^N
\end{aligned}$$

Weights

- weights are restricted to be non-negative and to sum to one,so synthetic controls are weighted averages of the units in the donor pool (control group)
- how the weights $\omega_2, \omega_3, ..., \omega_{J+1}$ should be chosen in practice.
- Some examples: $\omega_j = 1/J$ (equal weights).
- ω_j is the population in unit j (e.g., at the time of the intervention) as a fraction of the total population in the donor pool.
- Many other methods to estimate ω_j

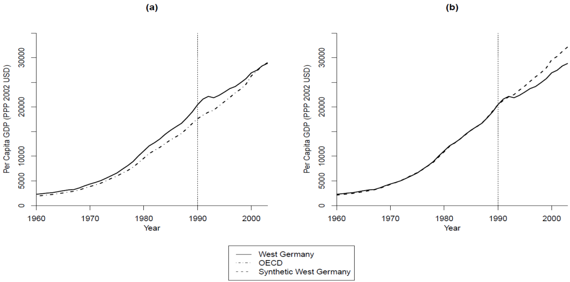
Weights

- Abadie et al. (2010) propose to choose the synthetic control, $W^* = \left(w_2^*, \ldots, w_{J+1}^*\right)$ that minimizes

$$\begin{aligned} \|X_1 - X_0 W\| &= \\ &= \left(\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \ldots - w_{J+1} X_{hJ+1}\right)^2 \right)^{1/2} \end{aligned}$$

The positive constants $v_1, \ldots, v_k$ reflect the relative importance of the synthetic control reproducing the values of each of the k predictors for the treated unit.

Weights


(a)
(b)
Per Capita GDP (PPP 2002 USD)
0
5000
10000
20000
30000
1960
1970
1980
1990
2000
Year
West Germany
OECD
Synthetic West Germany

Weights

Economic growth predictor means before the German reunification

	West Germany (1)	Synthetic West Germany (2)	OECD Sample (3)
GDP per-capita	15808.9	15802.24	13669.4
Trade openness	56.8	56.9	59.8
Inflation rate	2.6	3.5	7.6
Industry share	34.5	34.5	34.0
Schooling	55.5	55.2	38.7
Investment rate	27.0	27.0	25.9

Data requirements.

- Aggregate data on predictors and outcomesfor the unit exposed to the intervention of interest and a set of comparison units.
- Examples or these types of outcomes are state-level crime rates, country-level per-capita GDP, etc
- Sufficient pre-intervention information.
    - The credibility of a synthetic control estimator depends on its ability to track the trajectory of the outcome variable for the affected unit before the intervention.
- Suffcient post-intervention information.

Data requirements.

- Aggregate data on predictors and outcomesfor the unit exposed to the intervention of interest and a set of comparison units.
- Examples or these types of outcomes are state-level crime rates, country-level per-capita GDP, etc
- Sufficient pre-intervention information.
  - The credibility of a synthetic control estimator depends on its ability to track the trajectory of the outcome variable for the affected unit before the intervention.
- Suffcient post-intervention information.
  - Extensive post intervention information allows a more complete picture of the effects of the intervention

Thank you !