Lecture 10:
Synthetic Controls

March 25, 2020
Overview

Course Administration

Overview

Synthetic Control Set-up

Goal in Estimation

Convex Hull Required

Estimation

Examples
Course Administration

1. Hope everyone is doing ok. Let me know if you are having difficulties.
2. Today
   • Synthetic control: end by 7 pm
   • Then you do the in-class workshop virtually
3. Going forward
   • Lectures 11, 12, 13: I am available during this time for paper advice.
     • Share your screen! I’ll give Stata advice
     • We can talk through causality issues
     • Book me in advance
     • Will add an extra link near office hours with an additional scheduler
   • Lecture 14: I’ll pre-record a video on structural estimation and we can chat about it
4. GW moving to credit/no credit upon request for courses this semester
5. Additional changes to office hours as already noted
6. Anything else?
<table>
<thead>
<tr>
<th>Admin</th>
<th>Overview</th>
<th>Set-up</th>
<th>Goal</th>
<th>Convex Hull</th>
<th>Estimation</th>
<th>Examples</th>
</tr>
</thead>
</table>

**Overview**
When Synthetic Control?

- We would like to know the effect of a policy that happens in one (or a few) regions
- Why not diff-in-diff?
When Synthetic Control?

- We would like to know the effect of a policy that happens in one (or a few) regions
- Why not diff-in-diff?
  - small sample size $\rightarrow$ big standard errors
  - diff-in-diff requires that differences between treated and control are either
    - time-invariant, unit-specific measures or
    - time-varying in the same way for all units
- We can weaken these diff-in-diff assumptions by making a synthetic control
  - one comparison state that is a little of Michigan, a little of Illinois, no Wisconsin and a little Florida
- This doesn’t fix the small sample problem, but we use different inference methods
Set-up
• $t$ is time, $t \in \{1, \ldots, T_0, \ldots, T\}$.

• Treatment occurs after $T_0$.

• We look for effects starting in $T_0 + 1$
  • there are $T_0$ pre-intervention periods, $\{1, \ldots, T_0\}$
  • there are $T_1$ post-intervention periods, $\{T_0 + 1, \ldots, T\}$
  • total $T = T_0 + T_1$
Units in SC

• $i$ are observations, $i \in \{1, ..., J + 1\}$, 1 is treated, $\{2, 3, ..., J + 1\}$, or $J$ observations, are not. We call these $J$ observations the “donor pool”

• the donor pool should be
Units in SC

- $i$ are observations, $i \in \{1,..., J + 1\}$, 1 is treated, $\{2, 3, ..., J + 1\}$, or $J$ observations, are not. We call these $J$ observations the “donor pool”

- the donor pool should be
  - untreated (during observation period)
  - unaffected by treatment
  - should have no large, “idiosyncratic shocks,” to the outcome variable
  - “similar” to treated units to avoid interpolation bias (though it seems like the method should do this for you)
Outcomes for SC

- $Y_{it}^I$ outcome for treated
- $Y_{it}^N$ outcome for untreated
Outcomes for SC

- $Y_{it}^I$ outcome for treated
- $Y_{it}^N$ outcome for untreated
- we just observe $Y_{it}$
- we assume $Y_{it}^I = Y_{it}^N$ for any $t \leq T_0$
- $D_i = 1$ is ever treated, 0 is otherwise
- $Z_{it}$ are covariates
- define the effect of interest as $\alpha_{it} = Y_{it}^I - Y_{it}^N$
- what does this mean in words?
Outcomes for SC

- $Y_{it}$ outcome for treated
- $Y_{it}^N$ outcome for untreated
- we just observe $Y_{it}$
- we assume $Y_{it}^I = Y_{it}^N$ for any $t \leq T_0$
- $D_i = 1$ is ever treated, 0 is otherwise
- $Z_{it}$ are covariates
- define the effect of interest as $\alpha_{it} = Y_{it}^I - Y_{it}^N$
- what does this mean in words?
- note that this effect varies with time. How does this differ from a diff-in-diff?
Goal of estimation
Goal in estimation

- we want to find $\alpha_{it} = Y_{it}^I - Y_{it}^N$
- note that this is $\alpha_{it} = Y_{it} - Y_{it}^N$
- which of these do we know?
Goal in estimation

• we want to find $\alpha_{it} = Y_{it}^I - Y_{it}^N$
• note that this is $\alpha_{it} = Y_{it} - Y_{it}^N$
• which of these do we know?
• so the question is how to estimate $Y_{it}^N$
• intuition: approximate with a weighted average of non-treated units
• in math, $\hat{Y}_{it}^N = \sum_{j=2}^{J+1} w_j^* Y_{jt}$
Goal in estimation

- we want to find $\alpha_{it} = Y_{it}^I - Y_{it}^N$
- note that this is $\alpha_{it} = Y_{it} - Y_{it}^N$
- which of these do we know?
- so the question is how to estimate $Y_{it}^N$
- intuition: approximate with a weighted average of non-treated units
- in math, $\hat{Y}_{it}^N = \sum_{j=2}^{J+1} w_j^* Y_{jt}$

Trick is to find $w_j$
The Necessity of a Convex Hull
What is a Convex Hull?

- key requirement is that \( Y_{it} \) is in the convex hull of \( Y_{it}, i \neq j \)
- what is that?
  - in general, the convex hull of \( X \) is the “smallest convex set that contains \( X \)”
  - think of a set of three points \((x, y)\)
Convex Hull Example

Note: source is http://en.wikipedia.org/wiki/File:Convex_hull.png
What Has a Convex Hull?

• think of a treated observation where the donor pool would not form a convex hull.
What Has a Convex Hull?

- think of a treated observation where the donor pool would not form a convex hull.
  - impact of elections on growth in Afghanistan. there may be no obs that are in the convex hull for Afghanistan

- a sufficient condition for having donor pool observations in the convex hull is that the “number of pre-intervention periods is large relative to the scale of the transitory shocks.”

- assuming that the donor pool lies in the convex hull is equivalent to assuming $Y_{11t} - \sum_{j=2}^{J} w_j Y_{0jt} \equiv 0$ for $t < T_0$

- the convex hull assumption is sort of testable (maybe more on this later)
Estimation
What’s the Goal?

• in OLS, we minimize what?
What’s the Goal?

• in OLS, we minimize what? $\sum_{j=1}^{J} \epsilon_i^2$. In matrix language that $\epsilon' \epsilon$

• in this case, we choose weights to minimize the difference between the treated covariates and pre-treatment outcomes and the donor pool’s covariates and pre-treatment outcomes

• But remember that our optimal weights don’t have a time dimension.

• here we want to choose weights $W$ to minimize

$$||X_1 - X_0 W||$$

• $X$ contains both covariates $Z$ and pre-treatment outcomes $Y$
What we are minimizing

\[ X_1 = \begin{pmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1r} \\ Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1T_0} \end{pmatrix}, \quad X_0 = \begin{pmatrix} Z_{21} & \ldots & Z_{J+1,1} \\ Z_{22} & \ldots & Z_{J+1,2} \\ \vdots & \ddots & \vdots \\ Z_{2r} & \ldots & Z_{J+1,r} \\ Y_{21} & \ldots & Y_{J+1,1} \\ Y_{22} & \ldots & Y_{J+1,2} \\ \vdots & \ddots & \vdots \\ Y_{2T_0} & \ldots & Y_{J+1,T_0} \end{pmatrix}, \quad W = \begin{pmatrix} w_2 \\ w_3 \\ \vdots \\ w_J \end{pmatrix} \]
Weighting

- Implicitly equally weights all pre-treatment obs and covariates
- You can modify this, but each choice is a judgement call
- For example, combining the pre-treatment $Y$ into an average would down-weight them
- Note that $\|X_1 - X_0 W\|$ doesn’t give you one number – it gives you $r + T_0$ numbers.
- Final choice: how to weight those numbers when you add them up.
Example

Suppose that there is one covariate and two observations. The matrix looks like this:

\[
X_1 = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix}, \quad X_0 = \begin{pmatrix} 0 & 5 \\ 2 & 5 \\ 3 & 8 \end{pmatrix}, \quad W = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}
\]

\[
\|X_1 - X_0 W\| = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 0w_1 + 5w_2 \\ 2w_1 + 5w_2 \\ 3w_1 + 8w_2 \end{pmatrix} = \begin{pmatrix} 5 - (0w_1 + 5w_2) \\ 2 - (2w_1 + 5w_2) \\ 3 - (3w_1 + 8w_2) \end{pmatrix}
\]
And the Last Bit

- this outcome is a vector, and we have to decide how much we care about different parts of the diversion from the treated outcome.
- That “added-up number” is the mean squared error of the estimate. That is $\text{MSE} = \|X_1 - X_0 W\|_v$, where $v$ is yet another weighting matrix.
And the Last Bit

- this outcome is a vector, and we have to decide how much we care about different parts of the diversion from the treated outcome.

- That “added-up number” is the mean squared error of the estimate. That is $\text{MSE} = \|X_1 - X_0 W\|_\nu$, where $\nu$ is yet another weighting matrix

- How do you choose $\nu$? A variety of options
  - so that the pre-intervention difference in $Y$ is minimized
  - to minimize error in the final estimation (what they do in another, similar paper)
  - cross-validation in Germany paper:
    - find $W$ for the first half of the pre-treatment era
    - choose $\nu$ such that $\|X_1 - X_0 W\|_\nu$ is minimized in the second half of the pre-treatment period
    - if there are multiple possible $W$, you can see which one gives the lowest MSPE in the second pre-treatment period
And the Last Bit

- this outcome is a vector, and we have to decide how much we care about different parts of the diversion from the treated outcome.

- That “added-up number” is the mean squared error of the estimate. That is $\text{MSE} = ||X_1 - X_0 W||v$, where $v$ is yet another weighting matrix.

- How do you choose $v$? A variety of options
  - so that the pre-intervention difference in $Y$ is minimized
  - to minimize error in the final estimation (what they do in another, similar paper)
  - cross-validation in Germany paper:
    - find $W$ for the first half of the pre-treatment era
    - choose $v$ such that $||X_1 - X_0 W||v$ is minimized in the second half of the pre-treatment period
    - if there are multiple possible $W$, you can see which one gives the lowest MSPE in the second pre-treatment period

- Note that $||X_1 - X_0 W||v$ is the Mean Squared Prediction Error: MSPE
Re-capping Assumptions

- no effect of treatment on the untreated
- the treated unit would have had the untreated outcome in the absence of treatment
- treated observation is in the convex hull of the donor pool
Examples
Applying the Method

For Germany and breastfeeding

- What unit is treated?
- What are weights?
Applying the Method

For Germany and breastfeeding

- What unit is treated?
- What are weights? Germany paper is clear in Table 1
- How do we interpret main outcome tables?
  - Germany: Figures 1 and 2
  - Breastfeeding: Figures 1 to 4 (vertical line in wrong place)
- Other big-picture questions?