Synthetic Control Methods

David Sovich

Washington University in St. Louis
Some Motivation to Learn Synthetic Control

▶ “Arguably the most important innovation in the evaluation literature in the last fifteen years is the synthetic control method...this method builds on DD estimation, but uses arguably more attractive comparisons to get causal effects”

- Susan Athey and Guido Imbens (2016)

▶ Synthetic Control is already being used in finance, economics, and political science:

- One example: Acemoglu et. al (August - 2016, JFE)
Idea of Synthetic Control in a Visual

[Graph showing per-capita cigarette sales (in packs) over years, with a dotted line indicating the passage of Proposition 99 in 1990.]
So What is Synthetic Control?

- A method to evaluate the causal effect of shocks / policies

- SC builds upon the setting of the standard DD model, but makes two changes:
  1. Synthetic Control allows for time-varying individual-specific heterogeneity
  2. Synthetic Control takes a serious, data driven approach to forming counterfactuals / selecting the control group
Sounds Great, but what is the Catch?

- The benefits of Synthetic Control come with three costs:

  1. Large scale (asymptotic) inference cannot be conducted on Synthetic Control estimators
     - Instead, SC uses Permutation Methods to compute “standard errors”
  2. A key identifying assumption is somewhat ambiguous
  3. Computational time far exceeds that of DD estimators
Is There Anything Else to Tell me About Synthetic Control?

- Well, there are several other advantages SC possesses over DD

  1. Graphical Analysis - SC plots show how the ATT varies over time - also analyzes how well controls “fit” in pre-period

  2. Small Samples - Synthetic Control works well even when there is only one treated unit

  3. Heterogeneity Analysis - Synthetic Control can be used to estimate a separate ATT for each treated observation
Roadmap - A Lecture on Synthetic Control

1. Motivating Synthetic Control by Highlighting Weaknesses in Conventional DD

2. The Baseline Model for Synthetic Control

3. Getting Crafty: Moving to Multiple Treated Units and Multiple Treatments

4. Inference in Synthetic Control: An Area Still Under Construction

5. Conclusion
Program Evaluation - Weaknesses of DD
The Standard Program Evaluation Model

- Consider a simple two-period version of the potential outcomes framework of Rosenbaum and Rubin (1983)

- We are interested in the causal effect of a policy that occurs at time \( T = 1 \):
  \[
  y_{i,1}^1 - y_{i,1}^0
  \]

- Unobservability of counterfactuals and non-random assignment leads us to estimate the ATT:
  \[
  \mathbb{E} [y_{i,t}^1 - y_{i,t}^0 | \text{Treated} = 1, t = 1] ; \text{ATT}_0 = 0 \text{ By Definition}
  \]
The Difference-in-Differences Model

- Let treatment be assigned to state \( s = 1 \), but not state \( s = 0 \). We want to measure treatment’s effect on an individual outcome \( y \).

- The PO framework implies that the following holds \( \forall i \) and \( \forall t \):

\[
y_{i,t} = y_{i,t}^0 + (y_{i,t}^1 - y_{i,t}^0) \times D_{i,t}
\]

- Observed outcome \( y_{i,t} \) = outcome w/o treatment \( y_{i,t}^0 \) + treatment effect on i \( y_{i,t}^1 - y_{i,t}^0 \) \times binary policy indicator \( D_{i,t} \)

- Assuming that \( y_{i,1}^1 - y_{i,1}^0 = \alpha_1 \) in \( s = 1 \), we have \( \alpha_0 = 0 \) and

\[
y_{i,t} = y_{i,t}^0 + \alpha_t \times D_{i,t}
\]
The Difference-in-Differences Model

- The goal of policy evaluation is to estimate $\alpha_1$ - the effect of the policy on treated units. DD will make two assumptions to do so.

- **Assumption DD1 (Recoverability):** The counterfactual process, $y_{i,t}^0$, is determined by a variant of the following:

  $$y_{i,t}^0 = \delta_i + \delta_t + \epsilon_{i,t}$$

  - **Empiricist’s Note:** Industry-time fixed effects and observable control variables can be incorporated into the model as well.
Assumption DD1 implies that (see handout for derivations):

\[
\begin{align*}
\{ \mathbb{E} [y_{i,t} | s = 1, t = 1] - \mathbb{E} [y_{i,t} | s = 1, t = 0] \} - \\
- \{ \mathbb{E} [y_{i,t} | s = 0, t = 1] - \mathbb{E} [y_{i,t} | s = 0, t = 0] \} = \alpha_t + \\
\{ \Delta_t \mathbb{E} [\varepsilon_{i,t} | s = 1, t] \\
- \Delta_t \mathbb{E} [\varepsilon_{i,t} | s = 0, t] \}
\end{align*}
\]

Since \( \varepsilon_{i,t} = y_{i,t}^0 - \delta_i - \delta_t - \alpha_t D_{i,t} \), the above reduces to:

\[
\alpha_t + \{ \Delta_t \mathbb{E} [y_{i,t}^0 | s = 1, t] - \Delta_t \mathbb{E} [y_{i,t}^0 | s = 0, t] \}
\]
We want to recover $\alpha_t$ - so we make DD1 a little stronger

**Assumption DD2 (Parallel Trends):** In the absence of treatment, the $\Delta$ in the average values for the treated would match the controls:

$$\Delta_t \mathbb{E} [y_{i,t}^0 | s = 1, t] - \Delta_t \mathbb{E} [y_{i,t}^0 | s = 0, t] = 0$$

Best to think of in terms of “shocks” - e.g. the $\varepsilon_{i,t}$ notation - or differences in the counterfactual processes
The Key Weaknesses of Standard DD

▶ If assumptions DD1 and DD2 are met, then a panel regression recovers an unbiased estimate of the ATT:

\[ y_{i,t} = \delta_i + \delta_t + \beta D_{i,t} + \varepsilon_{i,t} \]

▶ DD1 ensures that diff. in expectations = ATT + Bias. DD2 ensures that bias = 0. Combined \( \implies \) reg. okay

▶ However, there are several circumstances in which the validity of these assumptions may be called into question
Violation of DD1 - Example

- What if the counterfactual process is more complicated?

\[ y_{i,t}^0 = \lambda_t \mu_i + \epsilon_{i,t} \]

and that differences in the loadings are correlated with treatment

- Then panel regression won’t recover an unbiased estimate of the ATT if there are differences in “shocks” across states that affect \( y \)
Violation of DD1 + DD2 - Example

✈️ If treatment is randomly assigned, then DD2 should always hold

✈️ We are never this lucky - so we must often select a control group for which we think DD1 + DD2 should hold

✈️ Most authors push this issue “under the rug” and openly put little thought into the control groups - e.g. All of Compustat

✈️ Even using matching techniques cannot solve this problem:

✈️ High dimension of covariate space makes it so perfect matching impossible - what if we could form “synthetic” matches though?
Motivation has been Built for Synthetic Control

- The key assumptions for DD to hold center largely around counterfactual selection:

  1. Treated and control units should be expected to follow similar factor models prior to treatment

  2. Moreover, they should not, on average, experience differential “shocks” in the post period

    - Structural break common to both groups is okay

- Synthetic Control places these issues up front and center and attempts to remove subjective control group formation
FOR MORE ON DD - SEE ASIDE
SYNTHETIC CONTROL MODEL
Single Treated Model - Setup

- The original SC model in Abadie et. al (2010) considers a single treated unit with multiple controls.

- Suppose we observe $J + 1$ states over $t = 1, \ldots, T$ periods, with state one being treated and states $\{2, \ldots, J + 1\}$ being unaffected.

- An “intervention” occurs at period $T_0 + 1$, $1 < T_0 + 1 < T$, that affects state one only, and leaves the other $J$ states unaffected.

- We aim to measure the impact of the policy on the treated state.
Single Treated Model - Setup

▶ Let $Y_{i,t}^N \equiv \text{outcome for state } i \text{ at time } t \text{ in the absence of intervention}$

▶ Let $Y_{i,t}^I \equiv \text{outcome for state } i \text{ that would be observed in time } t \text{ if unit } i \text{ was exposed to treatment}$

▶ We assume intervention has no effect on the outcome before implementation: $Y_{i,t}^N = Y_{i,t}^I$ for all $i$ and $t < T_0 + 1$

▶ **Note:** If there are anticipation effects, just redefine $T_0$ to some $\tau$

▶ We also assume the policy only affects the treated state
We aim to estimate the effect of intervention over time for the treated unit

\[ \alpha_1 = (\alpha_{1,T_0+1}, \ldots, \alpha_{1,T}) , \]

where for \( t > T_0 \):

\[ \alpha_{1,t} = Y^I_{1,t} - Y^N_{1,t} = \underbrace{Y_{1,t}}_{\text{observed}} - \underbrace{Y^N_{1,t}}_{\text{counterfactual}} \]

Therefore, similar to DD, we only need to construct the unobserved counterfactual
Single Treated Model - Identification Assumption One

▶ **Assumption SC1**: $Y_{i,t}^N$ follows a factor model for all $i$:

$$Y_{i,t}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{i,t}$$

(1)

where $\epsilon_{i,t}$ are zero mean shocks and:

▶ $\delta_t$ is an unobserved common factor - e.g. time FE

▶ $Z_i \in \mathbb{R}^r$ is a vector of observed covariates unaffected by intervention and $\theta_t' \in \mathbb{R}^r$ is a vector of loadings (parameters)

▶ $\lambda_t' \in \mathbb{R}^F$ is a vector of **common unobserved factors**, and $\mu_i \in \mathbb{R}^F$ is a vector of **unknown factor loadings**
An Aside: Thinking About Assumption SC1

- **Assumption SC1** is a more general version of **Assumption DD1**

  **Example:** Model of firm & time FEs given by \((\lambda_t = 1, \mu_i = \delta_i)\)

- **Assumption SC1** allows heterogenous responses to multiple unobserved factors \((\lambda_t \mu_i)\) and embeds time-trend models

- However, it implicitly assumes the number of factors \(|\lambda_t|\) are fixed over the period - e.g. no structural breaks
Assuming SC1, the basic idea behind SC estimation is as follows:

1. Reweight the control group so that a “synthetic” unit matches $Z_i$ and some pretreatment $Y_{i,t}$’s of the treated unit

2. As a result, $\mu_i$ will be automatically matched and the processes will be aligned

SC avoids “projecting” the FE out of the system - allowing for heterogenous responses to multiple unobserved factors
Let $W = (w_2, ..., w_{J+1})$ be a vector of weights with $w_j \geq 0 \ \forall j$

Idea is that each value of $W$ represents a potential synthetic control

For a given $W$, the outcome for a synthetic control at $t$ is:

$$Y_{W,t} = \sum_{j=2}^{J+1} w_j Y_{j,t}$$

$$= \delta_t + \theta_t \left( \sum_{j=2}^{J+1} w_j Z_j \right) + \lambda_t \left( \sum_{j=2}^{J+1} w_j \mu_j \right) + \left( \sum_{j=2}^{J+1} w_j \epsilon_{j,t} \right)$$
Theorem (SC Estimation): Suppose $\exists W^*$ such that the SC matches the treated unit in the pre-intervention period:

$$\sum_{j=2}^{J+1} w_j^* Y_{j,t} = Y_{1,t} \quad \forall t \in \{1, \ldots, T_0\}$$

and $\sum_{t=1}^{T_0} \lambda'_t \lambda_t$ is non-singular. Then for all $t > T_0$ we have

$$\mathbb{E} \left[ Y_{1,t}^N - \sum_{j=2}^{J+1} w_j^* Y_{j,t} \right] \rightarrow 0$$

as $T_0 \rightarrow \infty$ or if $T_0$ is large relative to $\varepsilon_{i,t}$ (see handout)
What Does This Even Mean?

- If the conditions are met, then the “synthetic control” associated with $W^*$ replicates the missing counterfactual

- An approximately unbiased estimator of $\alpha_{1,t}$ is then given by:

$$\hat{\alpha}_{1,t} = Y_{1,t} - \sum_{j=2}^{J+1} w^*_j Y_{j,t} = Y_{1,t} - Y_{W^*,t}$$

- e.g. the difference between the observed outcome and the synthetic control

- Beauty of SC is that even though the $\{\mu_i\}_i$ vectors are unobservable, fitting $\{Y, Z\}$ is sufficient to match the process
But Are the Assumptions Realistic? - Existence of $W^*$

- Note that the first part of the theorem can only hold if

$$\Lambda_1 \equiv (Y_1,1, \ldots, Y_1, T_0, Z'_1) \in \mathbb{R}^{1 \times (T_0 + r)}$$

belongs to the convex hull of $\{\Lambda_2, \ldots, \Lambda_{J+1}\}$

- This is a stronger requirement than $\{\Lambda_2, \ldots, \Lambda_{J+1}\}$ forming a basis for $\mathbb{R}^{1 \times (T_0 + r)}$ (e.g. $\text{Rank}(\{\Lambda_2, \ldots, \Lambda_{J+1}\}) = T_0 + r$) because the linear combination can only include positive weights

- If the treated unit has any outcome or $Z$ covariate “on the boundary”, then a perfect $W^*$ does not exist
But Are the Assumptions Realistic? - Existence of $W^*$

- Let us draw an example of when $W^*$ does and does not exist for $T_0 = 1$ and $r = 1$
But Are the Assumptions Realistic? - Existence of $W^*$

- The nice things about synthetic control is that we can actually check if this assumption holds.

- We can then exclude any treated observations for which we cannot exactly (or approximately) find a $W^*$.

- Moreover, Doudchenko and Imbens (2016) relax this assumption and propose an estimation strategy that allows for extrapolation.
But Are the Assumptions Realistic? - Existence of Factor Model

- **Assumption SC1** requires that treated and control units follow the same factor model over time.

- Abadie et. al (2010) suggest dropping observations for which we believe do not follow the same model.

- Note that DD makes a similar assumption, except that it is somewhat more restrictive.
But Are the Assumptions Realistic? - $T_0 \rightarrow \infty$

- The general consensus is that we should have a sufficiently large pre-treatment window to get an unbiased estimate of $Y^N_{i,t}$

- However, this assumption is not entirely needed if the size of $T_0$ is “large relative to $\varepsilon_{i,t}$” - e.g. the factor model fits well

- Moreover, the required size of $T_0$ varies by the type of process - AR(1) with time-varying coefficients only needs $T_0 = 1$
Empirically Estimating the Single Treated Model

- We estimate $\alpha_1$ by choosing a $W$ that forces the synthetic control to be as close to the treated unit.

- Currently, there are several different estimation schemes, each with distinct advantages and disadvantages.
  - These are discussed in detail in the handout.

- The Abadie et al. (2003, 2010, 2015) method is most commonly used and is implemented in the R package `synth`.
Application: California’s Proposition 99

- The following example is from Abadie et. al (2010) and Yiqing Xu’s (MIT) slides

- In 1988, California passed comprehensive tobacco control legislation:
  - Increased cigarette taxes by $0.25 per pack, funded anti-smoking media campaigns, and spurred clean-air ordinances

- We want to estimate the effect of the policy on smoking in California

- Other states that subsequently passed control programs excluded from donor pool
Cigarette Consumption: CA and the Rest of the US

![Graph showing per-capita cigarette sales in packs from 1970 to 2000. The graph compares California and the rest of the U.S. Passage of Proposition 99 is marked by a dotted line in 1990.]
Cigarette Consumption: CA and the Synthetic CA

![Graph showing per-capita cigarette sales in California and a synthetic California over the years from 1970 to 2000. The graph includes a note indicating the passage of Proposition 99 in 1986.]
Comparison of Synthetic Fit and Simple Average

<table>
<thead>
<tr>
<th>Variables</th>
<th>California Real</th>
<th>Synthetic</th>
<th>Average of 38 control states</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln(GDP per capita)</td>
<td>10.08</td>
<td>9.86</td>
<td>9.86</td>
</tr>
<tr>
<td>Percent aged 15-24</td>
<td>17.40</td>
<td>17.40</td>
<td>17.29</td>
</tr>
<tr>
<td>Retail price</td>
<td>89.42</td>
<td>89.41</td>
<td>87.27</td>
</tr>
<tr>
<td>Beer consumption per capita</td>
<td>24.28</td>
<td>24.20</td>
<td>23.75</td>
</tr>
<tr>
<td>Cigarette sales per capita 1988</td>
<td>90.10</td>
<td>91.62</td>
<td>114.20</td>
</tr>
<tr>
<td>Cigarette sales per capita 1980</td>
<td>120.20</td>
<td>120.43</td>
<td>136.58</td>
</tr>
<tr>
<td>Cigarette sales per capita 1975</td>
<td>127.10</td>
<td>126.99</td>
<td>132.81</td>
</tr>
</tbody>
</table>
Extending the Model to Multiple Treated Units
Extending the Baseline Model

- The baseline SC model in Abadie et. al (2010) only considers one treated unit

- But the model is extendable to multiple treated units
  - See Acemoglu et. al (2016), Cavallo et. al (2013), and Gobillion and Magnac (2016)
Extending the Baseline Model

- The main idea is to estimate a separate \( \alpha_i \) vector for each treated \( i \) using only control units when forming \( W_i^* \).

- We then calculate the ATT by “integrating” over the treated unit vectors

\[
\bar{\alpha} = \left\{ \bar{\alpha}_{T_0+1}, \ldots, \bar{\alpha}_T \right\} = \frac{1}{N_T} \sum_{i=1}^{N_T} \left\{ \hat{\alpha}_{i,T_0+1}, \ldots, \hat{\alpha}_{i,T} \right\}
\]

- Therefore, similar to the single treated case, we have dynamic estimate of the ATT.
Benefits of Synthetic Control with Multiple Treated Units

- Synthetic Control with multiple treated units possesses several advantages.

- We can easily examine the heterogeneity of treatment by looking how the $\hat{\alpha}_i$’s vary across observable dimensions.

- Moreover, we can visually inspect the fit of each $W_i^*$ and only keep treated units for which the assumptions likely hold.
Statistical Inference with Synthetic Control
Single Treated Unit - Abadie et. al (2010) Method

▸ Large-sample asymptotic inference is not possible with SC

▸ Abadie et. al (2010) suggest the use of Permutation Methods for inference

▸ Step One: Estimate a “placebo” treatment effect for each unit in the “donor pool” via SC methods

▸ Step Two: Calculate an empirical p-value for the effect estimated on the treatment unit

\[ p_1 = \frac{\sum_{j=2}^{J+1} 1\{\hat{\alpha}_j \geq \hat{\alpha}_1\}}{J} \]

▸ We can also calculate a \( p_{i,t} \) for each dynamic treatment effect instead of the average treatment effect
Smoking Gap for CA Versus Placebo Treatments
The Abadie method may provide poor inference when we include observations with “poor fitting” synthetic controls.

For example, when the SC fit is bad, we may get erroneous inferences - e.g. look at bottom gray line in Figure.

Abadie et. al (2012) propose looking at the Root Mean-Squared Prediction Error instead of $\hat{\alpha}_j$:

$$\text{RMSPE}_j = \frac{\sum_{t=T_0+1}^{T} (Y_{j,t} - \hat{Y}_{j,t}^N)^2}{\sum_{t=1}^{T_0} (Y_{j,t} - \hat{Y}_{j,t}^N)^2} = \frac{\text{Post} - \text{Period } "\text{Fit}"}{\text{Pre} - \text{Period } "\text{Fit}"}$$
Once we compute $RMPSE_j$ for each $j \in \{1, ..., J + 1\}$, we compare $RMPSE_1$ to the empirical placebo distribution

$$p_1 = \frac{\sum_{j=2}^{J+1} 1\{RMPSE_1 \geq RMPSE_j\}}{J}$$

Note 1 and $j$ are reversed.

Another method (Acemoglu) suggests removing observations with pre-period RMSPE’s of $\geq \sqrt{3} \times$ Average pre-period RMSPE.
California RMSPE Versus Donor Pool

(All 38 States in Donor Pool)

post/pre-Proposition 99 mean squared prediction error
Several recent methods have been developed for inference with multiple treated units.

These are also permutation methods, and they can be adapted to the single treated case as well.

Going forward let $N_T \equiv$ number of treated units and $N_C \equiv$ number of control units.
Multiple Treated Units - Acemoglu et. al (2016) Method

- Acemoglu et. al (2016) suggest constructing empirical confidence intervals among placebo policy groups

- Define the test statistic for the treatment group as

\[
\hat{\phi}(T) = \frac{\sum_{i \in \text{Treated}} \left( \frac{\sum_{t=T_0+1}^{T} Y_{i,t} - \hat{Y}_{i,t}^N}{\hat{\sigma}_i} \right)}{\sum_{i \in \text{Treated}} \hat{\sigma}_i^{-1}}
\]

where

\[
\hat{\sigma}_i = \sqrt{\frac{T_0}{\sum_{t=1}^{T_0} (Y_{i,t} - \hat{Y}_{i,t}^N)^2 / T_0}}
\]

- This method preserves the direction of the test and we can also remove outlier observations with \( \hat{\sigma}_i \geq \sqrt{3} \hat{\sigma}_{\text{Treated Average}} \)
Multiple Treated Units - Acemoglu et. al (2016) Method

- The test statistic is essentially a “fit weighted” estimate of the ATT (can change to be dynamic)

- We then compare the test statistic to an empirical placebo confidence interval

- Acemoglu et. al (2016) suggest forming 5,000 placebo treatment groups of size $N_T$ from the $N_C$ controls

- We then calculate $\hat{\phi}(Placebo)$ for each group, and compare $\hat{\phi}(T)$ to the distribution
Cavallo et. al (2013) provide a slightly different approach than Acemoglu et. al (2016)

They specifically make inference about the dynamics of the ATT, while Acemoglu et. al (2016) only reference it

Step One: compute the placebo dynamics for each \( j \in \text{Control Group} \)

\[
\hat{\alpha}_j = \{ \hat{\alpha}_{j,T_0+1}, \ldots, \hat{\alpha}_{j,T} \}
\]
Multiple Treated Units - Cavallo et. al (2013) Method

- Step Two: For each \( t \in \{ T_0 + 1, \ldots, T \} \), compute every possible average placebo treatment effect

  - Details: Let \( p \) denote a placebo group with size \( N_T \). The time \( t \) average placebo treatment effect is

  \[
  \hat{\alpha}_t^p = \frac{1}{N_T} \sum_{j \in p} \hat{\alpha}_{j,t}
  \]

  and we compute this effect for a total of \( K \equiv \binom{N_C}{N_T} \) placebo groups

- Step Three: For each \( t \in \{ T_0 + 1, \ldots, T \} \), compare \( \hat{\alpha}_t^{Treated} \) to the empirical distribution of placebo treatment effects and calculate

  \[
  p_t = \frac{\sum_{\ell=1}^{K} 1\{ \hat{\alpha}_{\ell,t} \geq \bar{\alpha}_t \}}{K}
  \]
Multiple Treated Units - ATT Over Time

**Figure 4. Adjusted Significance Levels for P99**

Lead Specific Significance Level (P-Values) for P99

Probability that this would occur by chance

Number of Years after a Large Disaster (Leads)
Practical Issues with Synthetic Control
Computation Time - Selection of Donor Pool

- The R Package \texttt{synth} implements a two-step optimization procedure to select the optimal $W^*$ for each treated $i$

- The optimization scheme suffers from a curse of dimensionality - lots of control units increase the number of potential $W$’s

- One way to reduce computation time is to restrict the control sample to units that are similar to treated units

  - Also helps fit the assumption that they follow the same factor process
Many of the inference methods require the use of a large number of placebo treatment groups

- Single treated problems only require placebo treated observations, multiple treated problems require placebo groups

Thus, it may be computationally infeasible to implement the Cavallo et al (2013) method with a large $N_T$ and $N_C$

- My advice: Limit yourself to a maximum of 5,000 placebo treatment groups
Conclusion
A Summary of Synthetic Control-DD

- There are two common problems that often in the context of DD studies:

  1. Given treated firms, how do we select the correct control firms?
     - Should hope for average expected change in treated to match the control. This issue is often pushed aside, leaving uncertainty about ability of control group to reproduce the counterfactual outcome trajectory

  2. DD estimators commonly only control for factors that are firm or time invariant.
     - Time varying confounders may pose identification concerns
A Summary of Synthetic Control-Pluses

▶ Synthetic Control tries to eliminate these concerns by:

1. Taking a data-driven approach to counterfactual formation - essentially forcing (in convex cases) the data to exhibit parallel trends in the pre-period

   ▶ Improvement upon even DD + Matching: often there does not exist a single unexposed unit that approximates the relevant characteristics of the treated unit

2. Nesting traditional additive fixed effects models by allowing a form of interactive fixed effects
A Summary of Synthetic Control-Minuses

- There are some drawbacks to the Synthetic Control method though:

1. Large scale (asymptotic) inference does not work with the Synthetic Control method
   - Need to use Permutation Methods to compute “standard errors”, and even here the approach and rejection rates vary across authors

2. For the Synthetic Control estimator to estimate the parameter of interest, the “training period must be large relative to the scale of the transitory shocks”
   - This is ambiguous, and it is unclear how to test necessary conditions of whether this holds in the data
ASIDE: NOTES ON DD
Some Additional Notes on DD

- **Note 1:** We can relax the assumption that the effect of treatment is the same ($\alpha_i$) for all treated units in the simple model. If we do this, we instead recover:

  $$\mathbb{E}[y_{i,t} | s = 1, t = 1]$$

- **Note 2:** The DD model can be generalized into a multiperiod setting. When we use a static estimator, we recover

  $$\int_{t \in \text{post}} \mathbb{E}[y_{i,t} | s = 1, t] \, dF(t)$$

  and in a dynamic setting we recover $\Gamma_\omega = \mathbb{E}[y_{i,t} | s = 1, \omega]$
Some Additional Notes: Dynamic DD

- Dynamic DD begins by first selecting a “reference year”
  
  - In the simple model this was clearly \( t = 0 \), but any pre-treatment period in which \( \alpha_t = 0 \) is expected should suffice

- Let this year be denoted as \(-T\), with time going from \{-T, \ldots, 0, \ldots, T\}, the PO framework implies

\[
y_{i,t} = y_{i,t}^0 + \sum_{\omega \neq -T}^{T} \Gamma_{\omega} D(i, \omega, t)
\]
Some Additional Notes: Dynamic DD

- Again, we notice that

\[
\{ \mathbb{E}[y_{i,t}|s=1,t=\omega] - \mathbb{E}[y_{i,t}|s=1,t=\omega] \} - \\
- \{ \mathbb{E}[y_{i,t}|s=0,t=-T] - \mathbb{E}[y_{i,t}|s=0,t=-T] \} = \Gamma \omega + \\
\{ \Delta_t \mathbb{E}[y_{i,t}^0 | s=1,\omega] \\
- \Delta_t \mathbb{E}[y_{i,t}^0 | s=0,-T] \}
\]

- And thus the **Parallel Trends** assumption must hold for each point in time
Some Additional Notes: Dynamic DD

- The dynamic DD setting is nice because it should be that $\forall t < 0, \Gamma_t = 0$ since $y_{i,t} = y_{i,t}^0 + 0$ before treatment occurs.

- Therefore, we can at least statistically reject that the assumption does not fail by examining the estimated $\Gamma_\omega$ coefficients for $\omega < 0$.

- Note also that with heterogeneous responses, we recover $\mathbb{E} [\Gamma_{i,\omega} | s = 1, t = \omega]$ from the regression.