Uncertainty Quantification in Synthetic Controls with Staggered Treatment Adoption

Matias D. Cattaneo, Princeton University
Yingjie Feng, Tsinghua University
Filippo Palomba, Princeton University
Rocío Titiunik, Princeton University

March 2024

https://nppackages.github.io/scpi/
Confounding by Unobservables

• In observational studies, treated and untreated units may not be directly comparable, even after adjusting for observed characteristics.
  
  • Key challenge: unobserved confounding variables.

• The difference-in-differences estimator eliminates time-invariant unobservable differences between treated and nontreated.
  
  • Identifying assumption relies on “parallel trends”.

• Synthetic control methods generalize the DiD framework by allowing for time-varying unobservable confounders.
  
  • Identifying assumptions depend on specific method considered.
  
  • Many “estimation” methods are available.
  
  • Few, often unprincipled, uncertainty quantification methods are available.
Synthetic control (SC) method

- Units of analysis: few aggregate entities, only one receives treatment.
- Causal effect of a treatment on a single unit.
- Popular in comparative case studies.
  - German reunification, California Tobacco Control, Compulsory Voting, etc.
- Key idea: Predict \( Y(0) \) of treated unit by linear combination of control units.
- Today: canonical single treatment unit case.
  - Paper/Software: multiple treatment units, staggered treatment adoption, more.

Our work focuses on uncertainty quantification for synthetic controls methods.

- Prediction Intervals for treatment effect on the treated.
- Two sources of uncertainty.
- Non-asymptotic coverage guarantees.
- Practical guidance.
Data Structure: Single Treatment Unit

Treatment: $D_{it} = 0, 1$  
Outcome: $Y_{it} = Y_{it}(1)D_{it} + Y_{it}(0)(1 - D_{it})$

\[
\begin{pmatrix}
✓ & ✓ & \cdots & ✓ \\
✓ & ✓ & \cdots & ✓ \\
\vdots & \vdots & \ddots & \vdots \\
✓ & ✓ & \cdots & ✓ \\
\end{pmatrix}
\begin{pmatrix}
✗ & \cdots & ✓ \\
✓ & \cdots & ✓ \\
\vdots & \vdots & \vdots \\
✓ & \cdots & ✓ \\
\end{pmatrix}
\rightarrow treated

\rightarrow untreated

\begin{align*}
Y(0) : ✓ &; \quad Y(1) : \times
\end{align*}
Synthetic Control: Basics

\[ Y_{it} = \begin{cases} 
Y_{it}(0) & \text{if } i = 2, \ldots, N + 1 \\
Y_{it}(0) & \text{if } i = 1 \text{ and } t \in \{1, 2, \ldots, T_0\} \\
Y_{it}(1) & \text{if } i = 1 \text{ and } t \in \{T_0 + 1, \ldots, T_0 + T_1\}.
\]

Treatment effect on the treated (random!)

\[ \tau_T = Y_{1T}(1) - Y_{1T}(0), \quad \text{for } T > T_0 \]

Find \( \{w_i\} \)

\[ \sum_{i=2}^{N+1} w_i Y_{it}(0) \approx Y_{1t}(0), \quad \text{for } t = 1, \cdots, T_0, \]

Hopefully,

\[ \sum_{i=2}^{N+1} w_i Y_{iT}(0) \approx Y_{1T}(0), \quad \text{for } T > T_0. \]

- Intuition: stable cross-sectional relation over time
**Synthetic Control: Basics**

\[ Y_{it} = \begin{cases} 
  Y_{it}(0) & \text{if } i = 2, \ldots N + 1 \\
  Y_{it}(0) & \text{if } i = 1 \text{ and } t \in \{1, 2, \ldots, T_0\} \\
  Y_{it}(1) & \text{if } i = 1 \text{ and } t \in \{T_0 + 1, \ldots, T_0 + T_1\} 
\end{cases} \]

Treatment effect on the treated \((\text{random!})\)

\[ \tau_T = Y_{1T}(1) - Y_{1T}(0), \quad \text{for } T > T_0 \]

Find \(\{w_i\}\)

\[ \sum_{i=2}^{N+1} w_i Y_{it}(0) \approx Y_{1t}(0), \quad \text{for } t = 1, \cdots, T_0, \]

Canonical SC

\[ \hat{w} = \arg \min_{w \in \mathcal{W}} \sum_{t=1}^{T_0} \left( Y_{1t}(0) - x_t' w \right)^2, \]

- \(\mathcal{W} = \{w \in \mathbb{R}^N : \sum_{i=2}^{N+1} w_i = 1, w_i \geq 0, \forall i \geq 2\} \quad x_t = (Y_{2t}(0), \cdots, Y_{N+1,t}(0))' \)
Synthetic Control: Basics

\[ Y_{it} = \begin{cases} 
Y_{it}(0) & \text{if } i = 2, \ldots, N + 1 \\
Y_{it}(0) & \text{if } i = 1 \text{ and } t \in \{1, 2, \ldots, T_0\} \\
Y_{it}(1) & \text{if } i = 1 \text{ and } t \in \{T_0 + 1, \ldots, T_0 + T_1\}.
\]

Treatment effect on the treated (random!)

\[ \tau_T = Y_{1T}(1) - Y_{1T}(0), \quad \text{for } T > T_0 \]

Find \( \{w_i\} \)

\[ \sum_{i=2}^{N+1} w_i Y_{it}(0) \approx Y_{1t}(0), \quad \text{for } t = 1, \ldots, T_0, \]

Canonical SC

\[ \hat{w} = \arg \min_{w \in \mathcal{W}} \sum_{t=1}^{T_0} \left( Y_{1t}(0) - x'_t w \right)^2, \]

- For a post-intervention \( T \geq T_0 \), the SC estimator is \( \hat{\tau}_T = Y_{1T} - \sum_{i=2}^{N+1} \hat{w}_i Y_{iT} \)
Example: 1990 German Reunification

Source: Abadie (2020)
<table>
<thead>
<tr>
<th></th>
<th>West Germany (1)</th>
<th>Synthetic West Germany (2)</th>
<th>OECD Sample (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP per-capita</td>
<td>15808.9</td>
<td>15802.24</td>
<td>13669.4</td>
</tr>
<tr>
<td>Trade openness</td>
<td>56.8</td>
<td>56.9</td>
<td>59.8</td>
</tr>
<tr>
<td>Inflation rate</td>
<td>2.6</td>
<td>3.5</td>
<td>7.6</td>
</tr>
<tr>
<td>Industry share</td>
<td>34.5</td>
<td>34.5</td>
<td>34.0</td>
</tr>
<tr>
<td>Schooling</td>
<td>55.5</td>
<td>55.2</td>
<td>38.7</td>
</tr>
<tr>
<td>Investment rate</td>
<td>27.0</td>
<td>27.0</td>
<td>25.9</td>
</tr>
</tbody>
</table>

Note: First column reports $Y_1(0)$, second column reports $\mathbf{x}_i' \hat{\mathbf{w}}$, and last column reports a simple average for the 16 OECD countries in the donor pool. GDP per capita, inflation rate, and trade openness are averages for 1981–1990. Industry share (of value added) is the average for 1981–1989. Schooling is the average for 1980 and 1985. Investment rate is averaged over 1980–1984.
Example: German Reunification

<table>
<thead>
<tr>
<th>country $j$</th>
<th>$[\hat{\omega}]_j$</th>
<th>country $j$</th>
<th>$[\hat{\omega}]_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0</td>
<td>Netherlands</td>
<td>0.10</td>
</tr>
<tr>
<td>Austria</td>
<td>0.42</td>
<td>New Zealand</td>
<td>0</td>
</tr>
<tr>
<td>Belgium</td>
<td>0</td>
<td>Norway</td>
<td>0</td>
</tr>
<tr>
<td>Denmark</td>
<td>0</td>
<td>Portugal</td>
<td>0</td>
</tr>
<tr>
<td>France</td>
<td>0</td>
<td>Spain</td>
<td>0</td>
</tr>
<tr>
<td>Greece</td>
<td>0</td>
<td>Switzerland</td>
<td>0.11</td>
</tr>
<tr>
<td>Italy</td>
<td>0</td>
<td>United Kingdom</td>
<td>0</td>
</tr>
<tr>
<td>Japan</td>
<td>0.16</td>
<td>United States</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Uncertainty Quantification

- The quantity of interest: $\tau_T = Y_{1T}(1) - Y_{1T}(0)$, $T > T_0$, the effect of the intervention on the treated unit.
  - Random variable unless potential outcomes are assumed fixed (Neyman, Fisher)!

- How do we quantify uncertainty? Popular answer: Permutation-Based Inference.
  - Abadie et al. (2010) propose inference approach for the synthetic control framework based on permutation methods.
  - A permutation distribution can be obtained by iteratively reassigning the treatment to the units in the donor pool and estimating “placebo effects” in each iteration.
  - The effect of the treatment on the unit affected by the intervention is deemed to be significant when its magnitude is extreme relative to the permutation distribution.
  - The permutation distribution is more informative than mechanically looking at $p$-values alone.
  - Depending on the number of units in the donor pool, conventional significance levels may be unrealistic or impossible.
  - Often, one sided inference is most relevant.
Example: California tobacco control program

![Graph showing per-capita cigarette sales in California and the rest of the U.S.](image)

- **Graph Details:**
  - **Y-axis:** Per-capita cigarette sales (in packs)
  - **X-axis:** Year (1970 to 2000)
  - **Key Events:**
    - Passage of Proposition 99

- **Legend:**
  - California
  - Rest of the U.S.
Example: California tobacco control program

Passage of Proposition 99

per-capita cigarette sales (in packs)

year
Example: California tobacco control program
Permutation-Based Inference

The graph shows the gap in per-capita cigarette sales (in packs) between California and control states over the years from 1970 to 2000. The passage of Proposition 99, indicated by the dotted line, is marked by a decrease in the gap, suggesting a significant impact on cigarette sales in California compared to control states.
Example: California tobacco control program
Permutation-Based Inference in Synthetic Controls

• This mode of inference reduces to classical randomization inference (Fisher, 1935) when the intervention is randomly assigned, a rather improbable setting.

• More generally, this mode of inference evaluates significance relative to a benchmark distribution for the assignment process, one that is implemented directly in the data.

• But, it is not based on an a known assignment mechanism.

  • It is often difficult to articulate the nature of the assignment mechanism or even the specific nature of a placebo intervention (France reunifies with whom?).

  • Even if a plausible assignment mechanism exists, estimation of the assignment mechanism is often hopeless, because many studies feature a single or a small number of treated units.

  • Potential outcomes are more naturally conceptualized as random variables in most applications (interval vs. external validity).
Prediction Intervals

Treatment effect on the treated

\[ \tau_T = Y_{1T}(1) - Y_{1T}(0), \quad \hat{\tau}_T = Y_{1T}(1) - \mathbf{x}_T' \mathbf{\hat{w}}, \quad T > T_0 \]

Prediction interval (PI) for random \( \tau_T \)

\[
\mathbb{P}\left\{ \mathbb{P}\left[ \tau_T \in \mathcal{I} \ \middle| \ \mathcal{H} \right] \geq 1 - \alpha \right\} \geq 1 - \pi
\]

- \( \mathcal{I} \): prediction interval
- \( \mathcal{H} \): conditioning \( \sigma \)-algebra
  - unconditional if \( \mathcal{H} \) is trivial
  - This paper: \( \{ \mathbf{x}_t : 1 \leq t \leq T \} \)
- \( 1 - \alpha \): conditional coverage prob., e.g., 95%
- \( \pi \): failure over \( \mathcal{H} \)
Prediction Intervals

Treatment effect on the treated

$$\tau_T = Y_{1T}(1) - Y_{1T}(0), \quad \hat{\tau}_T = Y_{1T}(1) - x'_T \hat{w}, \quad T > T_0$$

Prediction interval (PI) for random $\tau_T$

non-asymptotic: \[ \mathbb{P} \left\{ \mathbb{P} \left[ \tau_T \in I \mid \mathcal{H} \right] \geq 1 - \alpha \right\} \geq 1 - \pi \]

asymptotic: \[ \mathbb{P} \left[ \tau_T \in C \mid \mathcal{H} \right] \geq 1 - \alpha - o_{\mathbb{P}}(1) \]

• Link the two
  • Consider $\pi = o(1)$ (as $T_0 \to \infty$)
Two Sources of Uncertainty

Treatment effect on the treated

\[ \tau_T = Y_{1T}(1) - Y_{1T}(0), \quad \hat{\tau}_T = Y_{1T}(1) - x_T' \hat{w} \]

For a pseudo-true value \( w_0 \)

\[ Y_{1T}(0) = x_T' w_0 + u_T \]

A simple decomposition

\[ \hat{\tau}_T - \tau_T = \quad Y_{1T}(0) - x_T' \hat{w} \]
\[ = (x_T' w_0 + u_T) - x_T' \hat{w} \]
\[ = u_T - x_T' (\hat{w} - w_0) \]

• In-sample error: \( x_T' (\hat{w} - w_0) \)
• Out-of-sample error: \( u_T \)
• Non-asymptotically, both are important
Two Sources of Uncertainty

Treatment effect on the treated

\[ \tau_T = Y_{1T}(1) - Y_{1T}(0), \quad \hat{\tau}_T = Y_{1T}(1) - x'_T \hat{w} \]

For a pseudo-true value \( w_0 \)

\[ Y_{1T}(0) = x'_T w_0 + u_T \]

A simple decomposition

\[ \hat{\tau}_T - \tau_T = Y_{1T}(0) - x'_T \hat{w} \]
\[ = (x'_T w_0 + u_T) - x'_T \hat{w} \]
\[ = u_T - x'_T (\hat{w} - w_0) \]

- \( x'_T (\hat{w} - w_0) \): in-sample uncertainty coming from constructing SC weights using pre-treatment information.

- \( u_T \): out-of-sample uncertainty coming from misspecification plus any additional noise occurring at the post-treatment period \( T > T_0 \).
What is $w_0$?

Definition of $w_0$ relies on

- Constraint $\mathcal{W}$
- Assumptions on $\{Y_{1t}(0), x_t\}$
- Conditioning set $\mathcal{H}$ ($w_0$ may be random)

Examples

- $(Y_{1t}(0), x_t)$ stationary

$$w_0 = \arg \min_{w \in \mathcal{W}} \mathbb{E}\left[ \frac{1}{T_0} \sum_{t=1}^{T_0} (Y_{1t}(0) - x_t'w)^2 \bigg| \mathcal{H} \right]$$

  - (Constrained) best linear prediction

- $Y_{1t}(0), x_t \sim I(1)$ (integrated process)
  - Cointegration: $Y_{1t}(0) - x_t'w_0 \sim I(0)$
  - $(1, -w_0)$: cointegrating vector (unique given $w_0 \in \mathcal{W}$)
Prediction Intervals: Basic Construction

$$\hat{\tau}_T - \tau_T = u_T - \mathbf{x}_T'(\mathbf{\hat{w}} - w_0)$$

- **In-Sample Error**: with prob. $\geq 1 - \pi_1$ (over $\mathcal{H}$)

$$\mathbb{P}
\left[
M_{1,L} \leq \mathbf{x}_T'(w_0 - \mathbf{\hat{w}}) \leq M_{1,U} \mid \mathcal{H}
\right] \geq 1 - \alpha_1$$

- **Out-of-Sample Error**: with prob. $\geq 1 - \pi_2$ (over $\mathcal{H}$)

$$\mathbb{P}
\left[
M_{2,L} \leq u_T \leq M_{2,U} \mid \mathcal{H}
\right] \geq 1 - \alpha_2$$

**Prediction Interval** for $\tau_T$: with prob. $\geq 1 - \pi_1 - \pi_2$ (over $\mathcal{H}$)

$$\mathbb{P}
\left[
M_{1,L} + M_{2,L} \leq \hat{\tau}_T - \tau_T \leq M_{1,U} + M_{2,U} \mid \mathcal{H}
\right] \geq 1 - \alpha_1 - \alpha_2$$

- Conservative, but offer non-asymptotic probability guarantee
In-Sample Error: Stationary Case

\[
\sqrt{T_0}(\hat{w} - w_0) = \arg \min_{\delta \in \sqrt{T_0}(\mathcal{W}-w_0)} \ell(\delta), \quad \ell(\delta) = \delta' \left( \frac{1}{T_0} \sum_{t=1}^{T_0} x_t x_t' \right) \delta - 2 \left( \frac{1}{\sqrt{T_0}} \sum_{t=1}^{T_0} x_t' u_t \right) \delta
\]

- Assume \( \mathcal{W} \) is convex

- Possibly misspecified: \( \gamma := \mathbb{E}[\gamma | \mathcal{H}] \neq 0 \)

- By optimality of \( \hat{w} \) and \( w_0 \),

\[
\sqrt{T_0}(\hat{w} - w_0) \in \left\{ \delta \in \Delta : \delta' \hat{Q} \delta - 2(\hat{\gamma} - \gamma)' \delta \leq 0 \right\}, \quad \Delta = \sqrt{T_0}(\mathcal{W} - w_0)
\]

- \( \hat{Q} \) is fixed conditional on \( \mathcal{H} \)

- Approximate \( \hat{\gamma} - \gamma \)
In-Sample Error: Distributional Approximation

Goal: bound $x_T'(\hat{w} - w_0)$, e.g.,

$$\sqrt{T_0}x_T'(\hat{w} - w_0) \leq \sup_{\delta \in \mathcal{M}_{\hat{\gamma} - \gamma}} x_T' \delta$$

$$\mathcal{M}_{\hat{\gamma} - \gamma} = \left\{ \delta \in \Delta : \delta' Q \delta - 2(\hat{\gamma} - \gamma)' \delta \leq 0 \right\}$$

Reduce to distributional approximation of $\hat{\gamma} - \gamma$: for any $\kappa$,

$$\sup_{\delta \in \mathcal{M}_{\hat{\gamma} - \gamma}} x_T' \delta \leq \kappa \iff \hat{\gamma} - \gamma \in A_\kappa$$

• Normal approximation: Berry-Esseen bound

$$\mathbb{P}(\hat{\gamma} - \gamma \in A_\kappa) \approx \mathbb{P}(G \in A_\kappa), \quad G \sim \mathcal{N}(0, \nabla[\hat{\gamma}|\mathcal{H}])$$
In-Sample Error: Plug-in Approximation

$$\mathbb{P}(\sqrt{T_0}x'_T(\hat{w} - w_0) \leq \kappa) \geq \mathbb{P}(\hat{\gamma} - \gamma \in \mathcal{A}_\kappa) \approx \mathbb{P}(G \in \mathcal{A}_\kappa), \quad G \sim N(0, \nabla[\hat{\gamma} | \mathcal{H}])$$

$$\approx \mathbb{P}(\hat{G} \in \mathcal{A}_\kappa), \quad \hat{G} \sim N(0, \nabla[\hat{\gamma} | \mathcal{H}])$$

Basic approximation strategy:

$$\sup \left\{ x'_T \delta : \delta \in \Delta, \quad \delta' \hat{Q} \delta - 2(\hat{\gamma} - \gamma)' \delta \leq 0 \right\}$$

$$\Downarrow$$

$$\sup \left\{ x'_T \delta : \delta \in \Delta, \quad \delta' \hat{Q} \delta - 2G' \delta \leq 0 \right\}$$

$$\Downarrow$$

$$\sup \left\{ x'_T \delta : \delta \in \Delta^*, \quad \delta' \hat{Q} \delta - 2\hat{G}' \delta \leq 0 \right\}$$

- $\Delta^*$: locally equivalent to $\Delta = \sqrt{T_0}(W - w_0)$
- Thresholding
In-Sample Error: Non-Stationary Case

\[ T_0(\hat{w} - w_0) = \arg \min_{\delta \in T_0(w - w_0)} \ell(\delta), \quad \ell(\delta) = \delta' \left( \frac{1}{T_0} \sum_{t=1}^{T_0} x_t x_t' \right) \delta - 2 \left( \frac{1}{T_0} \sum_{t=1}^{T_0} x_t' u_t \right) \delta \]

Cointegrated system, e.g.,

\[ Y_{1t}(0) = x_t' w_0 + u_t, \quad x_t = x_{t-1} + \epsilon_t, \quad (u_t, \epsilon_t) \sim \text{i.i.d } I(0) \]

Again, bound \( T_0(\hat{w} - w_0) \) by simulating

\[ \sup \left\{ x_T' \delta : \delta \in \Delta^*, \delta' \hat{Q} \delta - 2 \hat{G}' \delta \leq 0 \right\}, \quad \hat{G} \sim \mathcal{N}(0, \hat{V}[\gamma|H]) \]

- \( \hat{Q} \) fixed conditional on \( H \)
- Probably \( \mathbb{E}[u_t|H] \neq 0 \)
- Non-stationarity affects the analysis of \( \hat{Q} \) and \( \mathbb{V}[\gamma|H] \)
Out-of-Sample Error

Three approaches:

- Concentration inequalities, e.g. subgaussian

\[ P\left( |u_T - \mathbb{E}[u_T | \mathcal{H}]| \geq \varpi_u | \mathcal{H} \right) \leq 2 \exp \left( - \frac{\varpi_u^2}{2 \sigma^2} \right) \]

- Location-scale model

\[ u_t = \mathbb{E}[u_t | \mathcal{H}] + (\nabla [u_t | \mathcal{H}])^{1/2} e_t, \quad \{e_t\} \perp \perp \mathcal{H} \]

- Quantile regression: model conditional quantiles of \( u_t \) given \( \mathcal{H} \)

Sensitivity analysis

- To cancel out the effect, how large does \( u_T \) need to be?
Example: Proposition 99 (California)
Example: Proposition 99 (California)

SC prediction with PI, in-sample error

Note: PI has $\geq 95\%$ coverage.
Example: Proposition 99 (California)

SC prediction with PI for $Y_{1T}(0)$, concentration-based

Note: PI has $\geq 90\%$ coverage.
Example: Proposition 99 (California)

SC prediction with PI for $Y_{1T}(0)$, local-scale model

Note: PI has $\geq 90\%$ coverage.
Example: Proposition 99 (California)

SC prediction with PI for $Y_{1T}(0)$, quantile-based

Note: PI has $\geq 90\%$ coverage.
General Form

Some extensions

- Match on multiple predictors
- Add more covariates to regression
- Other constraints \( \mathcal{W} \)

Goal: find \( \mathbf{w} = (w_2, \ldots, w_{N+1})' \) and \( \mathbf{r}_j = (r_{1,j}, \ldots, r_{k,j})' \)

\[
\begin{align*}
\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_M \end{bmatrix}_{T_0 \times 1} & \approx \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_M \end{bmatrix}_{T_0 \times N} \mathbf{w} + \begin{bmatrix} C_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{T_0 \times K} \mathbf{r}_j, & \text{for } j = 1, \ldots, M,
\end{align*}
\]

\[
(\hat{\mathbf{w}}, \hat{\mathbf{r}}) = \arg\min_{\mathbf{w} \in \mathcal{W}, \mathbf{r} \in \mathcal{R}} (A - B\mathbf{w} - C\mathbf{r})'(A - B\mathbf{w} - C\mathbf{r})
\]

\[
A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_M \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_M \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & 0 & \ldots & 0 \\ 0 & C_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & C_M \end{bmatrix}
\]
Conclusion

• Prediction Intervals for general SC methods.

• Conditional validity.

• Non-asymptotic probability guarantees.

• Two sources of uncertainty.

• Joint prediction intervals for staggered treatment adoptions (e.g., multiple treatment units).

• Principled tuning parameter selection.
   - Joint prediction intervals for staggered treatment adoptions (e.g., multiple treatment units treated at possibly different points in time).
   - Inference valid for locally quadratic constraints (e.g., ridge regression or L1-L2 regularization).
   - Principled tuning parameter selection and scalable, robust conic programming implementations.

   - Pointwise prediction intervals for one single treatment unit.
   - Inference valid for locally linear constraints.

   - [https://nppackages.github.io/scpi/](https://nppackages.github.io/scpi/)