

Synthetic Controls for Experimental Design

Alberto Abadie
MIT

Jinglong Zhao
Boston University

Workshop on Synthetic Control Methods
Center for Statistics and Machine Learning
Princeton University
June 3, 2022

Motivation

- ▶ Ride-sharing: what is the impact of a new driver-incentive pay program if deployed in the US?
- ▶ A user-level randomized control trial?
- ▶ Drivers in a city are randomized into the new program (treatment arm) or the status-quo (control arm)
- ▶ Drawbacks:
 - ▶ Fairness concerns: Drivers in different treatment arms obtain different compensations for the same jobs.
 - ▶ Interference: If drivers in the active treatment arm respond to higher incentives by working longer hours, they will effectively steal business from those in the control arm, which will result in biased experimental estimates.

Motivation

- ▶ The usual solution to interference is to randomize at the aggregate level, e.g., across cities.
- ▶ Assigning half of the cities to treatment and half to control?
- ▶ Drawbacks / business concerns:
 - ▶ Could be prohibitively expensive.
 - ▶ Could still raise substantial equity concerns.
 - ▶ Could create great disappointment among the drivers in the treated cities (half of the US drivers!) if the program is rolled back after experimentation.

Motivation

- ▶ Suppose instead that the new incentive pay treatment is applied and advertised as a pilot program in one city or in a few cities.
- ▶ Which city or cities should be treated?
- ▶ Which city or cities should be used as a comparison/control?
- ▶ This is a setting where randomization of treatment may create defective designs where:
 - ▶ The treated city/cities are non-representative of the entire set of cities of interest.
 - ▶ Treated and control cities are very different in their characteristics.

Synthetic controls for experimental design

- ▶ To address these challenges, we use the **synthetic control method as an experimental design** to select treated units in non-randomized experiments, as well as the untreated units to be used as a comparison group.
- ▶ We use the name **synthetic control designs** to refer to the resulting experimental designs.
- ▶ The choice of the treated unit(s) aims to accomplish two goals:
 - ▶ The features of the treated unit(s) should be representative of the features of an aggregate of interest, like the entire country.
 - ▶ The treated unit(s) should not be idiosyncratic in the sense that their features cannot be closely approximated by the units in the control arm.

Synthetic controls for experimental design

- ▶ In contrast to the observational case, in the experimental settings we have two synthetic units: one treated and one untreated.
- ▶ Related previous and ongoing works by Doudchenko and co-authors, and Jones and Barrows.
- ▶ Synthetic controls are widely used as observational and experimental designs by business analytics units (e.g., Amazon, Uber, Google, eBay, ...)

Conceptual framework

- ▶ T time periods and J units (cities in the ride-sharing example). T_0 pre-experimental periods.
- ▶ Using information available at T_0 , the analyst aims to select the set of units that will be administered treatment during the experimental periods, $T_0 + 1, T_0 + 2, \dots, T$.
- ▶ Potential outcomes:
 - Y_{jt}^I : potential outcome for unit j at time t under treatment.
 - Y_{jt}^N : potential outcome for unit j at time t under no treatment.
- ▶ Treatment effects:
$$Y_{jt}^I - Y_{jt}^N,$$
for $j = 1, \dots, J$ and $t = T_0 + 1, \dots, T$.
- ▶ Y_{jt} , the observed outcome, is Y_{jt}^I for treated units and Y_{jt}^N for untreated units.

Conceptual framework

- ▶ Suppose we aim to estimate the average treatment effect,

$$\tau_t = \sum_{j=1}^J f_j (Y_{jt}^I - Y_{jt}^N),$$

for $t = T_0 + 1, \dots, T$, where f_1, \dots, f_J are known weights (e.g., population share).

- ▶ An experimenter chooses $\mathbf{w} = (w_1, \dots, w_J)$ and $\mathbf{v} = (v_1, \dots, v_J)$, such that

$$w_j \geq 0, \quad v_j \geq 0,$$

$$w_j v_j = 0,$$

$$\sum_{j=1}^J w_j = 1, \quad \sum_{j=1}^J v_j = 1.$$

Conceptual framework

- ▶ Units with $w_j > 0$ are units that will be assigned to the intervention of interest from $T_0 + 1$ to T .
- ▶ Units with $w_j = 0$ constitute an untreated reservoir of potential control units (a “donor pool”). Among units with $w_j = 0$, those with $v_j > 0$ are used to estimate average outcomes under no intervention.
- ▶ A synthetic control estimator is

$$\tau_t(\mathbf{w}, \mathbf{v}) = \sum_{j=1}^J w_j Y_{jt} - \sum_{j=1}^J v_j Y_{jt},$$

where Y_{jt} are observed outcomes.

- ▶ How do we choose \mathbf{w}^* and \mathbf{v}^* ?

Conceptual framework

- ▶ The first goal of the experimenter is to choose w_1, \dots, w_J such that

$$\sum_{j=1}^J w_j Y_{jt}^I = \sum_{j=1}^J f_j Y_{jt}^I, \quad (1)$$

for $t = T_0 + 1, \dots, T$.

- ▶ The second goal of the experimenter is to choose v_1, \dots, v_J such that

$$\sum_{j=1}^J v_j Y_{jt}^N = \sum_{j=1}^J f_j Y_{jt}^N, \quad (2)$$

or, alternatively,

$$\sum_{j=1}^J v_j Y_{jt}^N = \sum_{j=1}^J w_j Y_{jt}^N. \quad (3)$$

Conceptual framework

- ▶ If (1) and (2) hold, then $\tau_t(\mathbf{w}, \mathbf{v})$ is equal to

$$\tau_t = \sum_{j=1}^J f_j(Y_{jt}^I - Y_{jt}^N),$$

which is the average treatment effect.

- ▶ If (3) holds, then $\tau_t(\mathbf{w}, \mathbf{v})$ is equal to

$$\tau_t^T = \sum_{j=1}^J w_j(Y_{jt}^I - Y_{jt}^N),$$

which is the average effect of the treatment on the treated (\mathbf{w} -weighted).

- ▶ We cannot directly fit potential outcomes because they are not directly observed. Instead, we will fit their predictors.

Estimation

- ▶ Define the estimation periods $\mathcal{E} \subseteq \{1, \dots, T_0\}$, $T_{\mathcal{E}} = |\mathcal{E}|$, and let $\mathbf{Y}_j^{\mathcal{E}}$ be the $(T_{\mathcal{E}} \times 1)$ vector of $T_{\mathcal{E}}$ pre-intervention outcomes for unit j .
- ▶ For any $j \in \{1, 2, \dots, J\}$, a vector of predictors, \mathbf{X}_j , is defined as

$$\mathbf{X}_j = \begin{pmatrix} \mathbf{Y}_j^{\mathcal{E}} \\ \mathbf{Z}_j \end{pmatrix},$$

where \mathbf{Z}_j are other pre-treatment covariates, aside from the outcomes, $\mathbf{Y}_j^{\mathcal{E}}$.

- ▶ The vector of predictors population averages is defined as

$$\bar{\mathbf{X}} = \sum_{j=1}^J f_j \mathbf{X}_j.$$

Estimation

$$\begin{aligned} \min_{\substack{w_1, \dots, w_J, \\ v_1, \dots, v_J}} & \left\| \bar{\mathbf{X}} - \sum_{j=1}^J w_j \mathbf{X}_j \right\|^2 + \left\| \bar{\mathbf{X}} - \sum_{j=1}^J v_j \mathbf{X}_j \right\|^2 \\ \text{s.t.} & \sum_{j=1}^J w_j = 1, \\ & \sum_{j=1}^J v_j = 1, \\ & w_j, v_j \geq 0, \quad j = 1, \dots, J, \\ & w_j v_j = 0, \quad j = 1, \dots, J, \\ & \underline{m} \leq \|\mathbf{w}\|_0 \leq \bar{m}, \end{aligned}$$

where $\|\mathbf{w}\|_0$ is the number of non-zero components of \mathbf{w} .

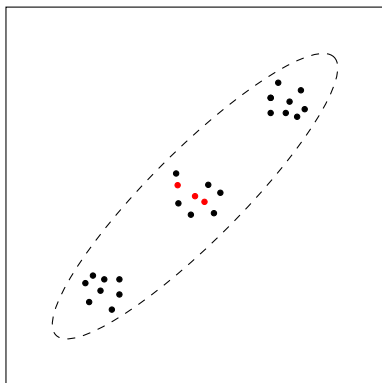
Estimation

$$\begin{aligned} \min_{\substack{w_j, \forall j=1,2,\dots,J, \\ v_{ij}, \forall i,j=1,2,\dots,J}} & \left\| \bar{\mathbf{X}} - \sum_{j=1}^J w_j \mathbf{X}_j \right\|^2 + \xi \sum_{j=1}^J w_j \left\| \mathbf{X}_j - \sum_{i=1}^J v_{ij} \mathbf{X}_i \right\|^2 \\ \text{s.t.} & \sum_{j=1}^J w_j = 1, \\ & w_j \geq 0, \quad j = 1, \dots, J, \\ & \sum_{i=1}^J v_{ij} = 1, \quad \forall j \in \mathcal{J}_{\mathbf{w}}, \\ & v_{ij} = 0, \quad \forall i \in \mathcal{J}_{\mathbf{w}}, j = 1, \dots, J \\ & v_{ij} \geq 0, \quad \forall j \in \mathcal{J}_{\mathbf{w}}, i = 1, \dots, J, \\ & v_{ij} = 0, \quad \forall j \notin \mathcal{J}_{\mathbf{w}}, i = 1, \dots, J, \\ & \underline{m} \leq \|\mathbf{w}\|_0 \leq \bar{m}. \end{aligned}$$

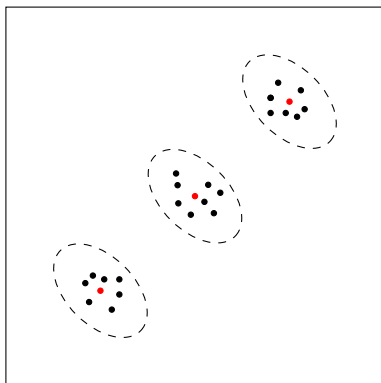
Then, make

$$v_j^* = \sum_{i=1}^J w_i^* v_{ji}^*.$$

Estimation



(a)



(b)

Panels (a) and (b) plot the same sample values of \mathbf{X}_j . Units assigned to treatment are drawn in red. In panel (a) we treat the entire sample as a single cluster. In panel (b) we divide the sample into three clusters and assign one unit in each cluster to the treatment.

Estimation

$$\min_{\substack{K, \mathcal{I}_1, \dots, \mathcal{I}_K, \\ w_j, \forall j=1, 2, \dots, J, \\ v_{ij}, \forall i, j=1, 2, \dots, J}} \sum_{k=1}^K \left(\sum_{j \in \mathcal{I}_k} f_j \right) \left\{ \left\| \bar{\mathbf{x}}_k - \sum_{j \in \mathcal{I}_k} w_j \mathbf{x}_j \right\|^2 + \xi \sum_{j \in \mathcal{I}_k} w_j \left\| \mathbf{x}_j - \sum_{i, j \in \mathcal{I}_k} v_{ij} \mathbf{x}_i \right\|^2 \right\}$$

$$\text{s.t. } \sum_{j \in \mathcal{I}_k} w_j = 1, \quad k = 1, \dots, K,$$

$$w_j \geq 0, \quad j = 1, \dots, J,$$

$$\sum_{i=1}^J v_{ij} = 1, \quad \forall j \in \mathcal{J}_{\mathbf{w}}$$

$$v_{ij} \geq 0, \quad \forall j \in \mathcal{J}_{\mathbf{w}}, i = 1, \dots, J,$$

$$v_{ij} = 0, \quad \forall j \notin \mathcal{J}_{\mathbf{w}}, i = 1, \dots, J,$$

$$v_{ij} = 0, \quad \forall i \in \mathcal{J}_{\mathbf{w}}, j = 1, \dots, J,$$

$$v_{ij} = 0, \quad k(i) \neq k(j),$$

$$\underline{m} \leq \|\mathbf{w}\|_0 \leq \bar{m},$$

where $k(j)$ is the cluster of observation j , and $\bar{\mathbf{x}}_k$ is a cluster mean.

Estimation

Assume that potential outcomes follow a linear factor model,

$$Y_{jt}^N = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_j + \boldsymbol{\lambda}_t \boldsymbol{\mu}_j + \epsilon_{jt},$$

$$Y_{jt}^I = v_t + \boldsymbol{\gamma}_t \mathbf{Z}_j + \boldsymbol{\eta}_t \boldsymbol{\mu}_j + \xi_{jt},$$

where \mathbf{Z}_j is a $(r \times 1)$ vector of observed covariates; $\boldsymbol{\theta}_t$ and $\boldsymbol{\gamma}_t$ are $(1 \times r)$ vectors of unknown parameters; $\boldsymbol{\mu}_j$ is a $(F \times 1)$ vector of unobserved covariates; $\boldsymbol{\lambda}_t$ and $\boldsymbol{\eta}_t$ are $(1 \times F)$ vectors of unknown parameters; ϵ_{jt} and ξ_{jt} are unobserved mean-zero random shocks.

Estimation

Assume also that, with probability one,

$$\sum_{j=1}^J w_j^* \mathbf{Z}_j = \sum_{j=1}^J f_j \mathbf{Z}_j, \quad \sum_{j=1}^J w_j^* Y_{jt} = \sum_{j=1}^J f_j Y_{jt}, \quad \forall t \in \mathcal{E},$$

and

$$\sum_{j=1}^J v_j^* \mathbf{Z}_j = \sum_{j=1}^J f_j \mathbf{Z}_j, \quad \sum_{j=1}^J v_j^* Y_{jt} = \sum_{j=1}^J f_j Y_{jt}, \quad \forall t \in \mathcal{E}.$$

Estimation

Under the assumptions above (and additional regularity conditions), we obtain

$$|E[\hat{\tau}_t - \tau_t]| \leq c \frac{\bar{\sigma}}{\sqrt{T_{\mathcal{E}}}},$$

where $\bar{\sigma}^2$ is the variance proxy of ε_{jt} (sub-Gaussian).

Inference

► Recall

$$Y_{jt}^N = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_j + \boldsymbol{\lambda}_t \boldsymbol{\mu}_j + \epsilon_{jt},$$

$$Y_{jt}^I = v_t + \boldsymbol{\gamma}_t \mathbf{Z}_j + \boldsymbol{\eta}_t \boldsymbol{\mu}_j + \xi_{jt}.$$

- Null hypothesis: For $t = T_0 + 1, \dots, T$, and $j = 1, \dots, J$,

$$Y_{jt}^I = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_j + \boldsymbol{\lambda}_t \boldsymbol{\mu}_j + \xi_{jt},$$

where ξ_{jt} has the same distribution as ϵ_{jt} .

- Under the null hypothesis, the distribution of Y_{jt}^I is the same as the distribution of Y_{jt}^N .

Inference

- ▶ Blank periods: $\mathcal{B} \subseteq \{1, \dots, T_0\} \setminus \mathcal{E}$, which comprise pre-intervention periods whose outcomes Y_{jt} have not been used to calculate \mathbf{w}^* or \mathbf{v}^*
- ▶ “Placebo” treatment effects estimated for the blank periods: for $t \in \mathcal{B}$,

$$\hat{u}_t = \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt}$$

- ▶ Post-intervention estimates of the treatment effects: for $t \in \{T_0 + 1, \dots, T\}$,

$$\hat{\tau}_t = \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt}$$

- ▶ Define the vector

$$\begin{aligned} \hat{\mathbf{r}} &= (\hat{r}_1, \dots, \hat{r}_{T-T_{\mathcal{E}}}) \\ &= (\hat{\tau}_{T_0+1}, \dots, \hat{\tau}_T, \hat{u}_{t_1}, \dots, \hat{u}_{t_{\mathcal{B}}}). \end{aligned}$$

Inference

- ▶ Permutation test: let Π be the set of all T_1 -combinations of $\{1, 2, \dots, T - T_{\mathcal{E}}\}$; for each $\pi \in \Pi$, let $\pi(i)$ be the i^{th} smallest value in π .
- ▶ Define the $(T_1 \times 1)$ -vector

$$\hat{\mathbf{e}}_{\pi} = (\hat{r}_{\pi(1)}, \hat{r}_{\pi(2)}, \dots, \hat{r}_{\pi(T_1)}).$$

- ▶ Test statistic:

$$S(\mathbf{e}_{\pi}) = \frac{1}{T_1} \sum_{t=1}^{T_1} |e_t|.$$

- ▶ p -value:

$$\hat{p} = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{1}\{S(\hat{\mathbf{e}}_{\pi}) \geq S(\hat{\mathbf{e}})\},$$

where $\hat{\mathbf{e}} = (\hat{r}_{T_0+1}, \dots, \hat{r}_T)$.

Inference

Under the assumptions above (and additional regularity conditions), we obtain that, for any $\alpha \in (0, 1]$,

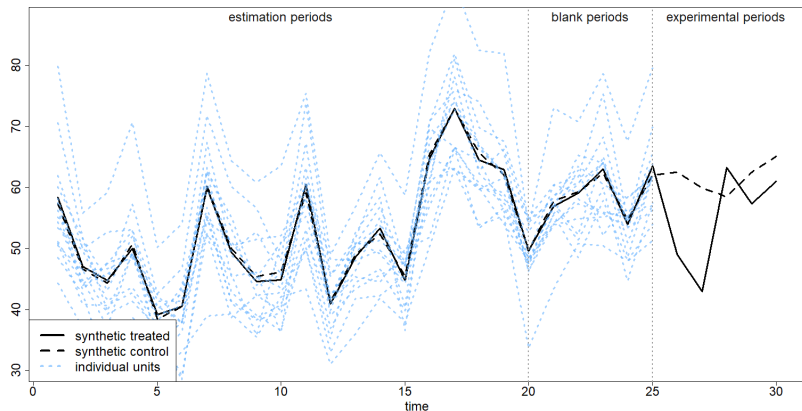
$$\alpha - \frac{1}{|\Pi|} \leq \Pr(\hat{p} \leq \alpha) \leq \alpha$$

with probability equal or greater than $1 - c/T_{\mathcal{E}}^{1/4}$.

Simulation setup

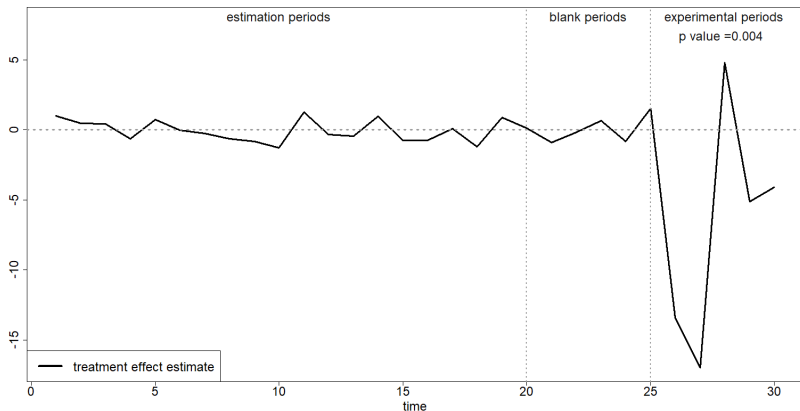
- ▶ $J = 15$ units, $r = 7$ observable covariates, $F = 11$ unobservable covariates
- ▶ $T = 30$ periods in total, $T_0 = 25$ pre-intervention periods, $T_1 = 5$ experimental periods
- ▶ Compute weights during the first $T_{\mathcal{E}} = 20$ periods, leave periods $t = 21, \dots, 25$ as blank periods
- ▶ δ_t and v_t as small-to-large re-arrangements of T i.i.d. Uniform $[0, 20]$ random variables
- ▶ \mathbf{Z}_j and $\boldsymbol{\mu}_j$ are random vectors of i.i.d. Uniform $[0, 1]$ random variables
- ▶ $\boldsymbol{\theta}_t$, $\boldsymbol{\gamma}_t$, $\boldsymbol{\lambda}_t$, and $\boldsymbol{\eta}_t$ are random vectors of i.i.d. Uniform $[0, 10]$ random variables
- ▶ ϵ_{jt} and ξ_{jt} are i.i.d. Normal $(0, 1)$ random variables

Results for one random simulation



The solid line represents the synthetic treated outcome (\mathbf{w}^* -weighted); the dashed line represents the synthetic control outcome (\mathbf{v}^* -weighted).

Results for one random simulation



This figure reports the difference between the synthetic treated and synthetic control outcomes. For the experimental periods, this is the treatment effect estimate.

Results over $B = 1000$ simulations

$$\frac{1}{B} \sum_{b=1}^B \tau_t^{(b)}$$

$t = 26$	$t = 27$	$t = 28$	$t = 29$	$t = 30$
-13.05	-10.92	-7.65	-5.17	-2.22

$$\frac{1}{B} \sum_{b=1}^B \hat{\tau}_t^{(b)}$$

		$t = 26$	$t = 27$	$t = 28$	$t = 29$	$t = 30$	MAE
<i>Unconstrained</i>		-13.06	-10.89	-7.63	-5.14	-2.22	0.85
<i>Constrained</i>	$\bar{m} = 1$	-12.99	-10.83	-7.74	-5.19	-2.34	2.84
	$\bar{m} = 2$	-12.99	-10.93	-7.58	-5.07	-2.17	1.66
	$\bar{m} = 3$	-13.04	-10.97	-7.58	-5.12	-2.13	1.22
	$\bar{m} = 4$	-13.06	-10.91	-7.59	-5.19	-2.19	1.02
	$\bar{m} = 5$	-13.06	-10.90	-7.63	-5.14	-2.20	0.92
	$\bar{m} = 6$	-13.07	-10.96	-7.64	-5.15	-2.22	0.88
	$\bar{m} = 7$	-13.06	-10.89	-7.63	-5.14	-2.21	0.85

Conclusions

- ▶ Experimental design methods have largely been concerned with settings where a large number of experimental units are randomly assigned to treatment and control.
- ▶ This focus on large samples and randomization has proven to be enormously useful in large classes of problems.
- ▶ However, it becomes inadequate when treating more than a few units is unfeasible, which is often the case in experimental studies with large aggregate units (e.g., cities).
- ▶ We have applied synthetic control techniques, widely used in observational studies, to the design of experiments when treatment can only be applied to a small number of experimental units.

Conclusions

- ▶ The synthetic control design optimizes jointly over the identities of the units assigned to the treatment and the control arms, and over the weights that determine the relative contribution of those units to reproduce the counterfactuals of interest.
- ▶ Corporate research units and academic investigators are often confronted with settings where interventions at the level of micro-units (i.e., customers, workers, or families) are unfeasible, unethical, impractical or ineffective.
- ▶ There is a wide range of potential applications of experimental design methods for large aggregate entities.

Synthetic Controls for Experimental Design

- ▶ Thanks for your attention!