Synthetic Controls for Experimental Design

Author names blinded for peer review

Consider the problem of a ride-sharing company choosing between two compensation plans for drivers (Doudchenko, Gilinson and Wernerfelt, n.d.; Jones and Barrows, 2019). The company can either keep the current compensation plan or adopt a new one with higher incentives. In order to estimate the effect of a change in compensation plans on profits, the company's data science unit designs and implements an experimental evaluation where the new plan is deployed at a small scale, say, in one of the local markets (cities) in the US. In this setting, a randomized control trial — or A/B test, where drivers in a local market are randomized into the new plan (active treatment arm) or the status-quo (control treatment arm) — is problematic. If drivers in the active treatment arm respond to higher incentives by working longer hours, they will effectively steal business from drivers in the control arm of the experiment, which will result in biased experimental estimates.

A possible approach to this problem is to assign an entire local market to treatment, and use the rest of the local markets, which remain under the current compensation plan during the experimental window, as potential comparison units. In this setting, using randomization to assign the active treatment allows ex-ante (i.e., pre-randomization) unbiased estimation of the effect of the active treatment. However, ex-post (i.e., post-randomization) biases can be large if, at baseline, the treated unit is different from the untreated units in the values of the features that affect the outcomes of interest. As in the ride-sharing example where there is only one treated local market, large biases may arise more generally in randomized studies when either the treatment arm or the control arm contains a small number of units, so randomized treatment assignment may not produce treated and control groups that are similar in their features.

To address these challenges, we propose the use of the synthetic control method (Abadie, Diamond and Hainmueller, 2010, Abadie and Gardeazabal, 2003) 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.[1]

[1] While we leave the "experimental" qualifier implicit in "synthetic control design", it should be noted that the

In our framework, the choice of the treated unit (or treated units, if multiple treated units are desired) aims to accomplish two goals. First, it is often useful to select the treated units such that their features are representative of the features of an aggregate of interest, like an entire country market. The treatment effect for the treated units selected in this way may more accurately reflect the effect of the treatment on the entire aggregate of interest. Second, the treated units should not be idiosyncratic in the sense that their features cannot be closely approximated by the units in the control arm. Otherwise, the reliability of the estimate of the effect on the treated unit may be questionable. We show how to achieve these two objectives, whenever they are possible to achieve, using synthetic control techniques.

While we are aware of the extensive use of synthetic control techniques for experimental design in business analytics units, especially in the technology companies,[2] the academic literature on this subject is at a nascent stage. There are, however, three publicly available studies that are connected to this article. To our knowledge, Doudchenko, Gilinson and Wernerfelt (n.d.) is the first (and only) publicly available study on the topic of experimental design with synthetic controls, and it is closely related to the present article. The focus of Doudchenko, Gilinson and Wernerfelt (n.d.) is on statistical power, which they calculate by simulation of the estimated effects of placebo interventions on historical (pre-experimental) data. That is, the selection of treated units is based on a measure of statistical power implied by the distribution of the placebo estimates for each unit. As a result, estimates based on the procedure in Doudchenko, Gilinson and Wernerfelt (n.d.) target the effect of the treatment for the unit or units that are most closely tracked in the placebo distribution. In the present article, we aim to take a different perspective on the problem of unit selection in experiments with synthetic controls; one that takes into account the extent to which different sets of treated and control units approximate an aggregate causal effect of interest.

In this paper, we propose various designs aimed to estimate average treatment effects, analyze

synthetic control designs proposed in this article differ from observational synthetic control designs (e.g., Abadie, Diamond and Hainmueller, 2010, Abadie and Gardeazabal, 2003, Doudchenko and Imbens, 2016), for which the identity of the treated unit(s) is taken as given.

[2]See, in particular, Jones and Barrows (2019), which also provides the basis for the ride-sharing example above.

the properties of such designs and the resulting estimators, and devise inferential methods to test a null hypothesis of no treatment effects. In addition, we report simulation results that demonstrate the applicability and computational feasibility of the methods proposed in this article.

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, impractical or ineffective (see, e.g., Duflo, Glennerster and Kremer, 2007, Jones and Barrows, 2019). There is, in consequence, a wide range of potential applications of experimental design methods for large aggregate entities, like the ones proposed in this article.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association*, 105(490): 493–505.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque County." *American Economic Review*, 93(1): 113–132.

Doudchenko, Nikolay, and Guido W Imbens. 2016. "Balancing, regression, difference-in-differences and synthetic control methods: A synthesis." National Bureau of Economic Research.

Doudchenko, Nikolay, Sean Gilinson, David; Taylor, and Nils Wernerfelt. n.d.. "Designing Experiments with Synthetic Controls."

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. "Using Randomization in Development Economics Research: A Toolkit." In . Vol. 4 of *Handbook of Development Economics*, , ed. T. Paul Schultz and John A. Strauss, Chapter 61, 3895–3962. Elsevier.

Jones, Nick, and Sam Barrows. 2019. "Synthetic Control and Alternatives to A/B Testing at Uber." *Presented at PyData Amsterdam 2019*, `https://youtu.be/j5DoJV5S2Ao`.