**Synthetic Control Method for Comparative Case Studies: FAQs**

Mithila Guha

The Synthetic Control Method (SCM) is a statistical method that was proposed by Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010, 2015) to identify the true causal impact of an aggregate-level intervention. This method is popular among multiple areas in the social sciences that use quasi-experimental data to estimate treatment effects. In this blog post, I will be answering different FAQ on the synthetic control method.

**1. What is the synthetic control method?**

The synthetic control method is a statistical approach used in comparative case studies that evaluate the true causal effect of intervention at an aggregate level (e.g. state, country, age group, etc.) by comparing the treatment group with a weighted combination of control groups.

**2. What is a synthetic control group?**

A synthetic control group can be defined as a weighted combination of several control groups from the donor pool to which the "treatment group is compared". By design of this method, the constructed synthetic control group needs to be approximately equal to the treatment group in the pre-treatment characteristics.

**3. What are the general assumptions used in the synthetic control method?**

Effective use of the SCM requires the following three general assumptions to hold:

i) Groups that have been exposed to a similar treatment are excluded from the pool of potential controls (the ‘donor pool’).

ii) The policy change has no effect before it is enacted.

iii) The treated group’s counterfactual outcome can be approximated by a fixed combination of donor controls.

**4. How are the comparison groups of a synthetic control method chosen?**

The synthetic control method allows researchers to systematically select comparison groups from a pool of potential candidates. The selection of donor groups and weights is determined by the pre-treatment outcome variable and the predictor variables that affect the outcome before the intervention.

**5. How are the predictor weights assigned to the synthetic control groups?**

According to Abadie, Diamond, and Hainmueller (2010), the predictor weights are assigned to minimize the mean squared prediction error (MSPE) of the predictor values in the pre-treatment years. The main target is to make the weighted combination of the selected control groups to be as similar as possible to the treated unit for the pre-treatment outcome variable and covariates. However, the weights must be non-negative and sum to one. In 2015, Abadie, Diamond, and Hainmueller expanded their suggestion by adding another systematic method to reduce overfitting through a form of cross-validation in which the pretreatment period is divided into a training period and a validation period.

**6. Do the weights assigned to the synthetic control groups change over time?**

The counterfactual synthetic control group in an SCM is built on a fixed combination of donor control groups. This means the weights do not change over time.

**7. How is the SCM model different than the 'difference-in-differences (DiD)' approach?**

The synthetic control method is a useful supplementary method to DiD, and it also combines elements from a statistical technique called 'matching'. The 'matching method' selects controls that have characteristics close to those of the treated group. The DiD method compares the average change over time in the outcome variable for the treatment group to that of the control group and calculates the effect of intervention at an aggregate level. The SCM combines these two methods. However, unlike the DID method, the SCM does not rely on parallel pre-implementation trends, and it also accounts for the effects of observed and unobserved time-varying confounders.

**8. How is the SCM model different than the causal impact approach?**

Both the SCM and the causal impact model use the help of control groups to construct a counterfactual of the treated group. As mentioned earlier, the SCM combines DiD techniques with elements from the matching method to construct the counterfactual. On the other hand, the causal impact method proposed by Brodersen et. al (2015) provides a fully Bayesian time-series estimate for the treatment effect by generalizing the DiD approach to the time-series setting. It is an improvised approach that uses Bayesian model averaging to construct the most appropriate synthetic control for modeling the counterfactual.

**9. When to use the synthetic control method?**

The SCM model is the most appropriate method to identify the true causal influence of a treatment for the following cases:

• When the intervention is done at an aggregate level, and there are a single or a few treated unit(s) with many potential control groups that are fixed in number.

• When the dataset consists of (long) longitudinal pre and post-treatment data.

• When the natural control groups are poor matches for the treatment group.

**10. What are the limitations of the synthetic control method?**

There are few limitations of the SCM model: i) The selection of the comparative units in a synthetic control method is not always well justified since there is currently no consensus on what constitutes a ‘good fit’ or how to judge the similarity.

ii) The credibility of the result relies on achieving a good pre-implementation fit for the outcome of interest between the treated unit and the synthetic control, which is difficult if the treated unit is an outlier.

iii) Another limitation of the SCM is that the traditional statistical inference is often inappropriate when there are a small number of treated and control units for large pre and post-treatment periods. Li (2019) resolved this difficulty by proposing a modified synthetic control method (MSC) to calculate the ATE, and by introducing a carefully designed subsampling method that provides valid confidence intervals for the ATE.

References:

Abadie, A., and J. Gardeazabal (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country", American Economic Review 93: 113–132.

Abadie, A., A. Diamond, and J. Hainmueller (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program", Journal of American Statistical Association 105: 493–505.

Abadie, A., A. Diamond, and J. Hainmueller (2015), "Comparative Politics and the Synthetic Control Method", American Journal of Political Science, Vol. 59, No. 2, April 2015, 495–510

Brodersen, K. H., F. Gallusser, J. Koehler, N. Remy, S. L. Scott (2015), "Inferring causal impact using Bayesian structural time-series models", Ann. Appl. Stat. 9, no. 1, 247--274. doi:10.1214/14-AOAS788. https://projecteuclid.org/euclid.aoas/1430226092

Li, K. T., (2019), "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods", Journal of the American Statistical Association, DOI: 10.1080/01621459.2019.1686986.