Sythetic prediction of treatment effects before they are implemented

Daniel J. Eck and Soheil Eshghi

February 10, 2020

Abstract

The synthetic control method (SCM) is a powerful methodology for evaluating the effects of treatments over time. In this article we reformulate SCM so that it can estimate treatment effects before they are implemented.

1   Introduction

Athey and Imbens [2017] state that SCM is "arguably the most important innovation in the policy evaluation literature in the last 15 years." We reformulate the synthetic control method [Abadie et al., 2010] as a tool to estimate multiple counterfactual with the aim of estimating what a treatment effect would be.

2   Background on SCM

What follows is adapted from Abadie et al. [2010]:

Let Y_{it}^N be the counterfactual outcome for subject i at time point t in the absence of treatment for subjects $i = 1, \ldots, n+1$, $t = 1, \ldots, T_i$. Let Y_{it}^I be the counterfactual outcome that would be observed for subject i at time t if subject i is exposed to the treatment. We let T_i^* be the time that the treatment is assigned to subject i where $T_i = \infty$ for subjects that do not receive the treatment. It is assumed that $Y_{it}^N = Y_{it}^I$ for all $t < T_i^*$.

Let $\alpha_{it} = Y_{it}^I - Y_{it}^N$ be the treatment effect for individual i at time period t, and let D_{it} be an indicator that is 1 if individual i is given the treatment and $t \geq T_i^*$. The observed outcome is

$$Y_{it} = Y_{it}^N + \alpha_{it} D_{it}.$$

In SCM the goal is to estimate $\alpha_{1T_1^*+1}, \ldots, \alpha_{1T_1}$ since subject 1 is the only individual who receives the treatment. To achieve this goal we need to estimate Y_{1t}^N for all $t > T_1^*$ since Y_{1t}^I is observed over this time period. Suppose that Y_{it}^N can be represented as a factor model,

$$Y_{it}^N = \delta_t + \theta_t' Z_i + \lambda_t' \mu_i + \varepsilon_{it},$$

where δ_t is an unknown common factor with constant factor loadings across units, $Z_i \in \mathbb{R}^p$ is a vector of observed covariates (not affected by the treatment), $\theta_t \in \mathbb{R}^p$ is a vector of unknown parameters, $\lambda_t \mathbb{R}^F$ is a vector of unobserved common factors, $\mu_i \in \mathbb{R}^F$ is a vector of unknown factor loadings, and the error terms ε_{it} are unobserved errors with $E(\varepsilon_{it}) = 0$.

Let $W = (w_2, \ldots, w_{n+1}) \in \mathbb{R}^n$ be a vector of weights that satisfy $\sum_{i=2}^{n+1} w_i = 1$ and $w_i \geq 0$ for all $i = 2, \ldots, n+1$. The value of the outcome variable for each synthetic control indexed by W is

$$\sum_{i=2}^{n+1} w_i Y_{it} = \delta_t + \theta_t' \sum_{i=2}^{n+1} w_i Z_i + \lambda_t' \sum_{i=2}^{n+1} w_i \mu_i + \sum_{i=2}^{n+1} w_i \varepsilon_{it},$$

and this synthetic control is then used to construct an estimate of the intervention effect,

$$\widehat{\alpha}_{1t} = Y_{1t} - \sum_{i=2}^{n+1} w_i Y_{it}.$$

The outcome variable of interest is observed for T_i periods, $t = 1, \ldots, T_i$, for the subject affected by the intervention, Y_{1t}, and the unaffected subjects, Y_{it}, $i = 2, \ldots, n+1$. Let $K_i \in \mathbb{R}^{T_i^*}$ define a linear combination of preintervention outcomes: $\bar{Y}_i^K = \sum_{t=1}^{T_i^*} k_t Y_{it}$. Consider M such linear combinations $K_1, \ldots, K_M$, and let $\mathbf{X}_1 = (Z_1', \bar{Y}_1^{K_1}, \ldots, \bar{Y}_1^{K_M})' \in \mathbb{R}^k$, $k = p + M$, be a vector of preintervention characteristics for the exposed subject. Similarly $\mathbf{X}_o \in \mathbb{R}^{k \times n}$ is a matrix containing the same variables for the unaffected subjects. That is, the ith column of $\mathbf{X}_o$ is $(Z_i', Y_i^{K_1}, \ldots, Y_i^{K_M})'$ (assuming that $T_i = T$ for all i). The vector W^* is chosen to minimize some distance, $\|\mathbf{X}_1 - \mathbf{X}_o W\|$. Abadie et al. [2010] proposed using $\|\mathbf{X}_1 - \mathbf{X}_o W\|_V = \sqrt{(\mathbf{X}_1 - \mathbf{X}_o W)' V (\mathbf{X}_1 - \mathbf{X}_o W)}$, where $V \in \mathbb{R}^{k \times k}$ is some symmetric and positive semidefinite matrix.

3 Synthetic prediction of treatment effect

Here we "reverse" SCM and use it as a methodology for constructing both the intervention and the nonintervention for subject 1 at future time periods. In this context we suppose that we have T_1^* preintervention observations of Y_{1t}^N and no observations of Y_{1t} for $t > T_1^*$. We aim to estimate α_{1t} for $t = T_1^* + 1, \ldots, T_1$. We achieve this by using SCM to estimate Y_{1t}^I and Y_{1t}^N for $t = T_1^* + 1, \ldots, T_1$.

We will now allow for two donor pools to estimate both counterfactuals Y_{1t}^N and Y_{1t}^I for $t > T_1^*$. We will let n_I and $n_N = n - n_I$ be the number of subjects in the donor pool used to estimate Y_{1t}^I and Y_{1t}^N respectively. We estimate Y_{1t}^I using data from the first donor pool. In this pool, we suppose that we have T_i^* observations of Y_{it}^N and $T_i - T_i^*$ postintervention observations of Y_{it}^I for subjects $i = 2, \ldots, n_I + 1$. We estimate Y_{1t}^N using data from the second donor pool. In this pool, we suppose that we have $T_1 + 1$ observations of Y_{it}^N for subjects $i = n_I + 2, \ldots, n+1$. Not all of the subjects are assumed to follow the same timelines as subject 1 in this context. We make the following assumption to ensure that the treatment assignments from subjects $2, \ldots, n+1$ do not interfere with the treatment effect on subject 1.

Assumption 1 (Independence) We assume that $(Y_{1t}^N, Y_{1t}^I, Z_1)$, $t = 1, \ldots, T_1^*$ are all independent of $(Y_{it}^N, Y_{it}^I, Z_i)$, $t = 1, \ldots, T_i$, $i = 2, \ldots, n+1$.

For now, we estimate the treatment effect $\alpha_{1T_1^*+1}$, the treatment effect for the first time period in which no counterfactuals Y_{1t}^I or Y_{1t}^N are observed.

Let $\mathbf{X}_1 = (Z_1', \bar{Y}_1^{K_1}, \ldots, \bar{Y}_1^{K_M})' \in \mathbb{R}^k$, $k = p + M$, be a vector of preintervention characteristics for subject 1. Let $\mathbf{X}_o^I \in \mathbb{R}^{k \times n_I}$ be a matrix containing the same variables for the n_I subjects belonging to the first donor pool. That is, the ith column of $\mathbf{X}_o^I$ is $(Z_i', Y_i^{K_1}, \ldots, Y_i^{K_M})'$ (assuming that $T_i = T$ for all $i = 2, \ldots, n_I + 1$). The vector W^{I*} is chosen so that

$$W^{I*} = \mathrm{argmin}_W \sqrt{(\mathbf{X}_1 - \mathbf{X}_o^I W)' V (\mathbf{X}_1 - \mathbf{X}_o^I W)},$$

where $V \in \mathbb{R}^{k \times k}$ is some symmetric and positive semidefinite matrix. Similarly, let $\mathbf{X}_o^N \in \mathbb{R}^{k \times n_N}$ be a matrix containing the same variables for the n_N subjects belonging to the second donor pool. That is, the ith column of $\mathbf{X}_o^N$ is $(Z_i', Y_i^{K_1}, \ldots, Y_i^{K_M})'$ (assuming that $T_i = T$ for all $i = n_I + 2, \ldots, n+1$). The vector W^{N^*} is chosen so that

$$W^{N^*} = \operatorname{argmin}_W \sqrt{(\mathbf{X}_1 - \mathbf{X}_o^N W)' V (\mathbf{X}_1 - \mathbf{X}_o^N W)}.$$

To ensure that both donor pools can reliably estimate the treatment effect α_{1,T_1^*+1} we assume that no treatment selection bias exists.

Assumption 2 (No selection bias) We assume that $(Y_{it}^N, Y_{it}^I) \perp\!\!\!\perp D_{it}$ for all $t = 1, \ldots, T_i$ and all subjects $i = 1, \ldots n+1$.

Under assumptions 1 and 2 we propose estimating α_{1,T_1^*+1} with

$$\widehat{\alpha}_{1,T_1^*+1} = \sum_{i=2}^{n_I+1} w_i^{I^*} Y_{iT_i^*+1} - \sum_{i=n_I+2}^{n+1} w_i^{N^*} Y_{iT_i^*+1}.$$

Notes

Taken from Soheil's notes (begin). Assume we have covariate-observation pairs $(y_i, \mathbf{x}_i)$ for n points, and that $\mathbf{x}_1$, the covariates of the intervened unit, lies within the convex hull of $(\mathbf{x}_i)_{i=1}^n$. One way to match covariates is to pick a vector $\mathbf{w}$ such that $\mathbf{w}'1 = 1$ and $\|\mathbf{x}_1 - \mathbf{Xw}\|$, where $\mathbf{X}$ is the matrix of available covariates (arranged in column form), is minimized.

One can incorporate a notion of locality by penalizing the use of covariates far from the covariates of interest, through incorporating the following term in either a constraint or as part of the objective: $\sum_{i=1}^n \mathbf{w}_i \|\mathbf{x}_1 - \mathbf{x}_i\|$.

This leads to the following optimization problem:

$$\begin{aligned}
\min_{\mathbf{w}} \quad & \|\mathbf{x}_1 - \mathbf{Xw}\| \\
\text{s.t.} \quad & \mathbf{w}'1 = 1, \\
& \sum_{i=1}^n \mathbf{w}_i \|\mathbf{x}_1 - \mathbf{x}_i\| \leq \delta, \\
& \operatorname{Var}(\mathbf{Yw}) \leq \gamma,
\end{aligned}$$

with δ and γ being rightsizing parameters.

We can simplify the above using the following assumptions:

… Once I have time I will write these out so that e can extract the fixed values mostly to one side.

The rest is kind of bad as is, to be honest, and hard to justify. I still think we should weigh down the ones that are pretty far from the expected value, but it's not clear why the variance condition above will not be enough and we have to do a prediction model for all $(n-1)-$ tuples. Any ideas?

The use of the transformed linear model (linking covariates to observations of outcomes) to create the counterfactual show our belief in the model specification. Under these conditions, the prediction of the model for the outcome at covariates $\mathbf{x}_i$ based on all other observed covariates $\mathbf{x}_{-i}$

(if it is within the convex hull) should closely match its observed outcome. Any discrepancy should make us less confident in the observation of the outcome (if we hold the model specification to be correct), so we would want y_i to play less of a role in the creation of the counterfactual in that scenario. Thus, we will put an upper limit on $\mathbf{w}_i$, the weight assigned to the covariate-observation pair by a function of the discrepancy between y_1 and the created synthetic prediction for it $\hat{y}_i$ from $\mathbf{x}_{-i}$, i.e.,

Taken from Soheil's notes (end).

References

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. 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, 2010.

Susan Athey and Guido W Imbens. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32, 2017.