STAT 8101: Applied Causality (Spring'22) Week 4: Feb 14, 2022 - Feb 20, 2022

Paper: Using Synthetic Controls

Han Lin hl3199@columbia.edu MS in Computer Science

This paper gives very good introduction to synthetic controls, which is very friendly to beginners in this topic. In section 3, the author mainly introduces the concept of synthetic control, which aims at approximating the trajectory that would have been observed in the absence of intervention. And it's nice to see the connection with counterfactuals.

In section 4, this paper also covers the relation of synthetic control with other alternative methods, which is very illuminative:

- Linear regression: compared with (linear) regression methods, synthetic control estimators preclude extrapolation by setting non-negative and unit sum constraints on weights, which will make the interpretation much easier: we could know explicitly the contribution from each comparison unit.
- Lasso: Both lasso and synthetic control could result in sparse weights. However, apart from the benefit from non-negative weights, the paper mentioned that the regularization used in lasso will "distort" the value of weight parameters, while the sparse weights from synthetic control are more interpretable because they are closer to the nature values.
- Panel data: in most panel data models, the effect of an intervention is restricted to be constant across post-intervention periods, while synthetic control could have different level of effects in each post-intervention period.

Then in section 5, the author discusses some practical requirements to use synthetic controls, and I'm mainly concerned about the selection of comparison groups.

The author mentions that the comparison group should contain some units not affected by interventions, as well as that units should have similar characteristics as the unit suffers from intervention. However, the author also mentions that the donor pool should not include units with large idiosyncratic shocks to the outcome variable, and there might be some potential spillover effect on units that are not directly targeted by it. Moreover, there might not exist a good combination of units in the comparison group that could approximate the treated unit well. So it seems that selection of a comparison group is kind of an art. I'm wondering are there any quantitative methods/criteria to evaluate the quality of a potential comparison group, or are there procedures/algorithms that could select a subset as comparison gropu from a larger pool?

The case where there exist multiple treated units seems also quite interesting, the author points out several recent advances in this field. I will read more about this later.