What Is the Value Added by Uing Causal Machine Learning Methods in a Welfare Experiment Evaluation?
by
Anthony Strittmatter

Kuan Xu
Dalhousie University, Halifax, Canada

October 29, 2019

The contributions

This study investigates whether CML methods add value compared to conventional CATE estimators by re-evaluating Connecticut's Jobs First welfare experiment.
Previous studies show that it is hard to tackle the effect heterogeneity of Jobs First by means of CATEs.
This study shows evidence that CML methods can provide support for the theoretical labor supply predictions.
It also provides reasons why some conventional CATE estimators fail and discuss the limitations of CML methods.

Hypotheses

H1 The Jobs First program has positive earnings effects for some individuals with zero earnings under AFDC.

H2 There is a mix of positive and negative earnings effects in the group of individuals with positive earnings below the FPL under AFDC.

H3 Jobs First has non-positive earnings effects for individuals with earnings above the FPL under AFDC.

Data

The Connecticut Department of Social Services required the MDRC to conduct a randomized controlled trial to evaluate the Jobs First program. Experimental participants were single-parent welfare applicants and recipients who lived in Manchester or New Haven. Between January 1996 and February 1997, 4,803 experimental participants were randomly assigned to either the AFDC (control group) or Jobs First (treatment group) programs. This is based on the experiment data!

Variables

- Outcome: earnings
- Covariates: baseline (elapsed quarters), decent (13 including age, education, children, welfare history, etc.), and kitchen sink (68 variables affecting heterogeneity)

Methodology

- Although it is an experiment, only one potential outcome is observable for the members of either control group or treatment group. All participants have their characteristics which can be conditioned upon. Treatment is randomly assigned.
- Local conditional model: local arbitrary grouping;
- CML—double machine learning:

$$outcome = f(covaraites, selection(covariates))$$

- ML estimators
  - Tree estimator—regression trees select how to partition the sample;
  - General random forest estimator—forecasts are assembled from many trees using different sub-samples and different sub-sets of covariates;

Findings

CML methods can add value to a Jobs First evaluation in the sense that they can provide evidence supporting the theoretical labor supply predictions.
However, this strategy works only when the CML methods incorporate many important heterogeneity variables and allow for continuously distributed treatment effects.

Issues worth considering

- Strengthen the links between the theory and predictions;
- H2 is not a conventional hypothesis (not in the sense of a one- or two-tailed test);
- CML helps in model specification:
  - Englarge the information set for CATE or QTE;
  - Avoid misspecifications in grouping, discretization, or omitted variables;
  - Searching for functional forms via variations across covariates and subsamples;
- Is "incorporating many important heterogeniety variable" a limitation?
- Bootstrap inference versus proportions.