ECON 293 Homework 2: Commentary

Jack Blundell, Spring 2017

In this commentary I discuss results and include some key figures. Many further figures and all code are provided in the attached .html file. I worked with Luis Armona on the code. This commentary is written individually.

As in the previous homework, we use data from "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment", by Dean Karlan and John List (AER 2007). In this experiment, two-thirds of recipients of a charity solicitation letter receive some kind of match-donation treatment, whereas the remaining control group receives the letter alone, with no matching promise. We use the same censoring rule as in homework 1 to eliminate observations, emulating an observational study for the first part of this homework. For reference, the 'true' ATE, that which is estimated on the full sample in which treatment is allocated randomly, is 0.1536.

1 Observational study

[See 'causaltreebriefintro'. very useful.]

This first set of methods is applied to our censored data, emulating observational data with non-random treatment assignment.

1.1 Propensity forest

Our first method is propensity forest from package causalTree. In this method, we use a random forest where in each tree we split the sample into leaves based on propensity score, then estimate the treatment effect within each leaf. This means that we only use the treatment assignment indicator W_i to make splits. This is then a fairly natural extension of propensity score matching using machine learning methods. We can summarise the estimates with the average treatment effect, calculated by averaging over all estimated treatment effects. Note that when applying propensity forests there is no need for a hold-out sample for honest estimation. The estimated treatment effect is calculated over the same sample as that on which the model is fit. In doing so, we get an estimate of 0.3535. Relative to our estimates in homework 1, we see that this method actually performs relatively poorly.

1.2 Gradient forest

Gradient forests (Athey, Tibshirani and Wager 2016) allows the estimation of heterogeneous parameters at a particular point in covariate space by using 'neighborhoods' of observations in a training set. Many existing methods to do this suffer from the curse of dimensionality and are extremely computationally intensive. Gradient forests solves many of these computational problems, using an adaptive weighting function derived from random forests. When fitting trees, observations are labeled with the gradient of the estimating equation. Next, as is standard when using trees, observations are split into leaves and parameters of interest are estimated within leaves. Importantly, it has been shown that this method delivers consistent, asymptotically normal estimates. Treatment effects are estimated honestly, meaning that they are estimated on a seperate sub-sample.

Using gradient forest, we obtain an estimated ATE of 0.3285.

As an alternative, we first residualize our data. To do so, we use a random forest to predict our outcome and treatment variables, based on covariates. We then subtract these predicted values to obtain residuals. With these residuals of Y and W we then run gradient forest just as above. We obtain an estimated ATE of 0.2275.

Comparing to our other estimates, we see that the residualized gradient forest does best out of the methods tried in this homework here so far. However, we discovered in homework 1 that OLS with controls in fact performed very well relative to the ML methods. This result has not changed. It is perhaps unsurprising that our methods here do not represent an improvement in estimating the ATE. These methods are after all mostly designed to detect heterogeneity in treatment effects rather than efficiently estimating the ATE in observational settings.

2 Randomized trial

In this section I discuss results on the full sample, in which treatment is allocated randomly.

2.1 LASSO

Here, I use LASSO to select significant interaction terms between treatment and covariates. I estimate a LASSO on training sample A, including on the right hand side treatment, all covariates and interactions between treatment and covariates. I use cross validation to select significant terms. Since the standard approach does not pick up any heterogeneity, I set the penalty factor for one interaction which I know to be important in the full dataset to zero, so that it is artificially chosen by LASSO.

I then run post-OLS on the selected interaction, the covariate in question ('perbush') and treatment. I perform this regression on sample A, B, C, as well as the union of B and C. My estimates are as follows....

[INSERT TABLE]

2.2 Honest causal tree

Now I fit an honest causal tree using the causalTree package. This works similarly to most tree methods in machine learning, with the major difference being that splits are made according to estimated treatment effects. After a tree is built, I can prune the tree using cross validation. We are left with a group of leaves containing observations with similar estimated treatment effects. 'Honesty' here means that we use a different sample for estimation within leaves to that which we used for making the splits and cross validation.

I fit my model using sample A and estimate treatment effects within leaves seperately on samples B and C. I obtain

2.3 Causal forest

Next I use causal forest, also from the causalTree package. For this an all remaining parts of the homework, I pool samples A and B into a single training set. In a causal forest, an ensemble of trees are built. For each of these trees, a random sample is taken from the training data. To predict, an average value over all fitted tree predictions is used for each individual observation.

2.4 Causal forest (Gradient forest)

Next I use the causal.forest command from the gradient.forest package. This is just as described in part I above. The only difference is that because treatment is randomized here, we do not residualize treatment.

2.5 Discussion

Table X summarizes the average treatment effects estimated across this section. I also give the MSE against Y^*, which here is the outcome variable for each observation divided by the propensity score if treated, and one minus the propensity score if not treated. Since we have a random experiment, the propensity score is just the proportion of treated observations in the sample.

[WHICH METHODS GIVE MORE HETEROGENEITY?] [HOW DO THE AVERAGE ESTIMATES DIFFER?]