PROJECT: ELV 781

Submitted by

Aravind J (2017EET2297), Vinay Kyatham (2017EET2307)

Part1:

Replicating the Results in Table 2 and 6 of Angrist and Evans

Calculate the means and standard deviations for Y_i, X_i, W_i

Part2:

Replicating the Results of columns 1, 4 and 7 of Table 7 of Angrist and Evans

OLS estimation of Y_i on X's and W

Part3:

Run causal forest for each Yi

Part 3A:
plain decision tree on treatment 0 and treatment 1 data separately
Part 3B:
decision tree using TOT
Part 3C:
decision tree using \hat{tau}

How to run the code:

1. Download and place the 'AngEv98.zip' in the working directory of the project

2. Run data_loader.py this will convert the sas data into numpy data that is used by the main program

3. Run main.py. This will print out the replicated results for tables 2, 6 and 7.

This will also save the graphical representation of the trees in the folder "save_trees" in the current directory.

Part1:

TABLE2- DESCRIPTIVE STATISTICS, WOMENAGED 21-35 WITH 2 OR MORECHILDREN (1980 PUMS)

	All women	Married women
Children ever born	2.5610691	2.516374
More than 2 children	0.411164	0.387715
Boy 1st	0.5001088	0.51714
Boy 2nd	0.51096	0.51437
Two boys	0.27341	0.26706
Two girls	0.2416	0.24055
Same sex	0.5153895	0.50162

TABLE 6-OLS ESTIMATES OF MORE THAN2 CHILDREN EQUATIONS 1980 PUMS

	All women(1)	All women(2)	Married women(4)	Married women(5)
Boy Ist	-	0.00803102829843138	-	0.01108670
Boy 2nd	-	0.008980280860919186	-	-0.0116788
Same sex	0.05964402230979977	0.0612525867500454	0.06653557572430279	0.069168

Part2:

TABLE7-OLS ESTIMATES OF LABOR-SUPPLY MODELSUSING 1980 CENSUSDATA

	All women	Married women	Husbands of married women
Worked for pay	-0.1737863292529353	-0.16188512684270634	-0.009132659718460507
Weeks worked	-8.787188198603152	-7.825325181335769	-0.9255253683950512
Hours/week	-6.442167957126217	-5.746952586164105	0.03561086574148787
Labor income	-3616.619190781292	-3013.1406801912913	-1471.5655427565146
ln(Family income)	-0.16584063009883906	-0.1535747466066865	_

Part3:

In this section we trained decision trees for three different methods specified.

We use a 90:10 train:test split to find the best depth for the decision tree. We allow a maximum depth of 15. We found it adequate for the given task.

The graphical representations of the trees are saved in the folder "saved_trees" in the working directory. For the three different cases, three folders are created to store the trees generated.