MSE 228 HW7 Part 1

bkohen, Wai Yam Gordon Tsai

March 2023

1   a

Our dataset is from an experiment conducted in Brazil to study the effects of high school financial education. The outcome was measured as the student's score in a financial knowledge test. About 50 percent of the students are selected for treatment at random, where the treatment is receiving financial education during high school. This corresponds to an RCT setting.

The data is collected from schools through three survey rounds consisting of baseline (August 2010), follow-up 1 (December 2010) and follow-up 2 (December 2011). In addition, complementary administrative data on school graduation and dropout rates was compiled for the study period.

The dataset consists of 17000 samples and 16 variables. The dataset was collected through questionnaires given to students before financial education (in the case of treated students). The financial proficiency score was collected by a test after the education period. You can see a description of all variables in the page below.

Panel A. School-level variables
Number of students in school (2008 data)
Number of teachers in school (2008 data)
Grade-level dropout rate (2009 data)
Grade-level passing rate (2009 data)

Panel B. Student background characteristics
Student is female
Mother attended secondary school
Father attended secondary school
Student has failed at least one school year
Student's family receives *Bolsa Familia* cash
    transfer
Student has computer with Internet at home
Student has some form of income
Student is unemployed

Panel C. Student financial characteristics
Financial proficiency score
Saves money for future purchases
Intention to save index
Makes a list of expenses every month
Negotiates prices or payment methods
Financial autonomy index

2   b
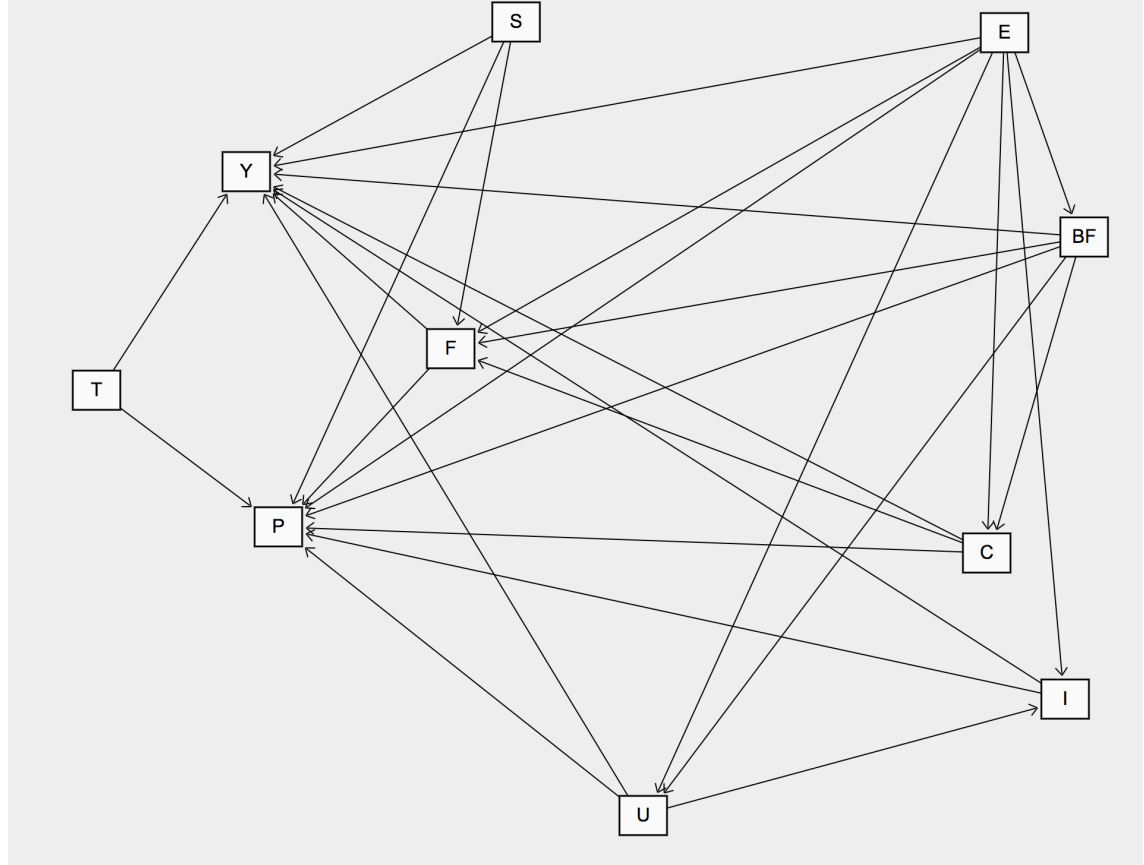
T is our treatment variable, which is whether a student was included in the financial education program.

Y is the outcome variable, which is the financial proficiency score of the student that is measured after the treatment.

P includes all post treatment variables (except the outcome). This includes financial characteristics that are measured with a test.

The rest of the variables are all pre-treatment. S is school wide variables. F is whether the student has failed in the past. E is whether the parents have secondary education. BF is whether the family receives cash through the Bolsa Famila government program for low-income families. C is whether the family has a computer. I is whether the student has some source of income. U is whether the student is unemployed.

3   c

1. We will run an OLS of Y (financial proficiency) on D (financial education), D*X, X, where X is all the pre-treatment variables. As seen in our DAG, D is independent of all variables in X, so we will identify the ATE with this model. We will exclude all post-treatment variables.

2. We will run a Double Lasso for the predictive effect of D in the BLP of y using D, D*X, X

3. We will run several meta learner methods such as S-learner, T-learner, Cross-learner, and DR-learner. We will do this for the gender variable, because it is interesting to see whether the treatment effect changes based on the gender of the student. This could have implications on whether the education programme was gender-biased. We will also do it for the dropout-rate of the school, to see if the impact of financial education is different based on the characteristics of a school. For example, schools with higher dropout rates could see a smaller increase in the financial proficiency score because students are less attentive on average. However, they could also see a larger increase because higher dropout rates could be correlated with lower income areas where financial education would make a bigger difference.

4   d

1. For OLS we will report confidence intervals and heteroscedasticity robust standard errors for our treatment variable and all other pre-treatment variables. We will also report the p-value for the treatment variable.

2. For Double Lasso and Elastic Net, we will report confidence intervals and standard errors. For Double Lasso, we will report which variables had non zero coefficients.

3. We will report out of sample DR-score, calibration score, p-values on BLP using the predicted CATE and Qini curve for CATE estimators.