EC709 PS2. Treatment Effects

Iván Fernández-Val

Boston University

Due on October 3

Question 1: Effect of 401(k) on Asset Accumulation In this question you are going to estimate the causal effect of 401(k) eligibility and participation on accumulation of assets. The file `sipp1991.dta` contains information of 9,915 households from the Survey of Income and Program Participation (SIPP) for 1991 on the following variables:

```
Contains data from sipp1991.dta
  obs:         9,915
 vars:            14                          20 Jun 2013 14:08
------------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------------------------
nifa            float   %9.0g                 Net non-401(k) financial assets
net_tfa         float   %9.0g                 Net total financial assets
tw              float   %9.0g                 Total wealth
age             byte    %9.0g                 Age of the head of the household
inc             float   %9.0g                 Household income
fsize           byte    %9.0g                 Household size
educ            byte    %9.0g                 Years of education of the head of the household
db              byte    %9.0g                 Defined benefit pension status indicator
marr            byte    %9.0g                 Married indicator
twoearn         byte    %9.0g                 Two-earner status indicator
e401            byte    %9.0g                 401(k) eligibility
p401            byte    %9.0g                 401(k) participation
pira            byte    %9.0g                 IRA participation indicator
hown            byte    %9.0g                 House ownership indicator
------------------------------------------------------------------------------------------------
Sorted by: e401
```

1. Verify that
$$\mathrm{E}\left[\frac{(1-D)Y}{1-P(X)}\right] = \mathrm{E}[\mathrm{E}[Y \mid X, D=0]],$$
where $P(X) = \mathrm{P}(D=1 \mid X)$ is the propensity score.

2. Estimate the ATE of 401(k) *eligibility* on net total financial assets using the nonparametric regression, propensity score reweighting and double robust estimators with a

low dimensional specification for the controls such as the one seen in class. Do your estimates have a causal interpretation? Define the causal parameter and provide its identification conditions in the context of this application.

[Hint: you can use OLS to estimate conditional expectations and logit to estimate the propensity score.]

3. Estimate the LATE of 401(k) *participation* on net total financial assets with 401(k) eligibility as an instrument using the nonparametric regression, propensity score reweighting and double robust estimators with a low dimensional specification for the controls such as the one seen in class. Do your estimates have a causal interpretation? Define the causal parameter and provide its identification conditions in the context of this application.

4. Estimate the LATE of 401(k) *participation* on net total financial assets with 401(k) eligibility as an instrument using the double robust estimator with a high dimensional specification for the controls such as the one seen in class.

   [Hint: you can use Lasso to estimate regression functions of continuous outcomes and logit Lasso to estimate regression functions of binary outcomes. The package `hdm` implements Lasso and logit Lasso regressions in R.]

5. Extra credit: use bootstrap to compute the standard errors of your estimates in parts 2 and 3.

   [Hint: the `R` package `boot` is very useful to implement the bootstrap. It has functionality for parallel computing.]

Question 2: Head Start (Ludwig and Miller, 2007) Ludwig and Miller (2007) estimated the effect of the Head Start program on health and schooling outcomes using a regression discontinuity design.[1] The discontinuity is induced by the program design that targeted just the 300 poorest counties in the country setting a cutoff of 59.1984 in the poverty rate for eligibility into the program. The file `headstart.dta` contains information of 2,810 counties on the following variables:

```
Contains data from headstart.dta
  obs:         2,810
 vars:            13                         29 May 2021 13:52
-----------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
```

[1]Ludwig, J., and Miller, D. L. (2007): "Does Head Start improve children's life chances? Evidence from a regression discontinuity design," *Quarterly Journal of Economics*, 122, 159–208.

```
------------------------------------------------------------------------------------------------
povrate60       float   %9.0g                 County Poverty Rate 1960, HS cutoff = 59.1984
mort_age59_re~S float   %9.0g                 Mortality, Ages 5-9, HS related causes, 1973-1983
mort_age25plu.. float   %9.0g                 Mortality, Ages 25+, HS related causes, 1973-1983
mort_a~s_postHS float   %9.0g                 Mortality, Ages 25+, Injuries, 1973-1983
census1960_pop  long    %12.0g                Census 1960: county population
census1960_pc~7 float   %9.0g                 Census 1960: % attending school, age 14-17
census1960_pc~4 float   %9.0g                 Census 1960: % attending school, age 5-34
census1960_pc~s float   %9.0g                 Census 1960: % high-school or more, age 25+
census1960_po~7 long    %12.0g                Census 1960: population, age 14-17
census1960_po~4 long    %12.0g                Census 1960: population, age 5-34
census1960_po~s long    %12.0g                Census 1960: population, age 25+
census1960_pc~n float   %9.0g                 Census 1990: % urban population
census1960_pc~k float   %9.0g                 Census 1990: % black population
------------------------------------------------------------------------------------------------
Sorted by:
```

Read Ludwig and Miller (2007) and reanalyze the data using state of the art methods.

[Hint: Another useful reading is Cattaneo, Titiunik and Vazquez-Bare (2017): "Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality," *Journal of Policy Analysis and Management* 36(3): 643-681.]