Assignment 4

Due November 15th by 11:59pm PT. Download your notebook with your code and responses as a PDF via HTML and submit to Gradescope. Remember to tag the sections in your assignment that correspond to the questions on Gradescope.

For this problem set, we are going to use data from:

Lee, K., Miguel, E., & Wolfram, C. (2020). Experimental evidence on the economics of rural electrification. Journal of Political Economy, 128(4), 1523-1565.

We already used this dataset in Small Assignment 2, but we will revisit our results now that we have a better understanding of RCTs and the potential outcome framework. We have now the following variables of interest:

- TakeUp_i, a binary variable equal to 1 if a household connects to the electricity grid and 0 otherwise
- Treatment_i, a binary equal to 1 if a household is in any of the treatment groups (described below) and 0 otherwise
- Price_i,the effective price to pay for the connection to the grid (discounted from the value of the subsidy that households received)
- Earnings, which denotes monthly earnings of the household
- HhSize, household size
- Female, a binary variable equal to 1 if the household head is female
- Age, the age of the household head
- Finschool, a binary variable equal to 1 if the household head finished secondary school
- Spouse, a binary variable equal to 1 if the household head is married
- NotFarmer, a binary variable equal to 1 if the household head is not a farmer
- Employed, a binary variable equal to 1 if the household head is employed
- Bank, a binary variable equal to 1 if the household head has a bank account
- OwnLand, the acres of land owned by the household

Note that all the variables except TakeUp_i, Treatment_i and Price_i are measured at baseline, i.e. before the start of the program.

This model is estimated with a random sample of 2275 households in rural Kenya. Subsidies are randomly allocated as such:

1. High subsidy: 380 unconnected households in 25 communities are offered a $\$398 (100\%)$ subsidy, resulting in an effective price of $\$0$.

2. Medium subsidy: 379 unconnected households in 25 communities are offered a $\$227 (57\%)$ subsidy, resulting in an effective price of $\$171$.
3. Low subsidy: 380 unconnected households in 25 communities are offered a $\$114 (29\%)$ subsidy, resulting in an effective price of $\$284$.
4. Control group: 1150 unconnected households in 75 communities are not offered any subsidy and face regular connection price of $\$398$.

Question 1

Suppose the researchers did not conduct a randomized controlled trial. Instead they surveyed households in rural Kenya and compared the take up of electrification in places where the government decided to subsidize electrification (non-randomly) with the take up in places where farmers faced the regular connection price. Would this comparison allow you to estimate the effect of subsidies on the take up of electrification in rural areas?

Explain why or why not using both words and potential outcomes notation. How would the expression for the estimated impact of subsidies on tree cover differ if cash incentives were randomized? For this part, assume that you group all subsidies into one treatment group, denoted $T_i = 1$ if the household is selected to receive a subsidy, and $T_i = 0$ if the household is allocated to the control group.

Question 2

a) Before running any regressions, show how you can obtain the average treatment effect (ATE) of the subsidies (still grouped) on the probability of taking up the connection to the grid. Then write down the regression you could use to estimate the ATE.

b) Read in the data from `dta_LMW_PS4.dta` with the `read_dta()` function from the `haven` package. Compute the ATE as you first wrote it down. Then, estimate the regression you just wrote down and test whether the ATE is statistically significant at the 1% significance level.

In [ ]:

Question 3

a) Now let's check to make sure our sample is balanced across treatment and control. For each of the household demographics variables, conduct a t-test against the null that they are on average equal between treatment and control.

In [ ]:

b) Do you conclude that the sample is well-balanced? Are you surprised or concerned by any of the results? If so, what could you do to address your concerns?

Question 4

a) What happens if you control for additional covariates in this regression? What (if any) advantages are there to controlling for these variables? What (if any) disadvantages are there?

b) Run the regression controlling for all nine of these variables. Interpret how your results change, if at all.

In [ ]:

Question 5

You think that households heads with above average income might be more incentivized from the subsidy in terms of taking up the connection to the grid because they might be more able to afford it. Oppositely, you might think that richer households are less senzitive to a subsidy since they might need it less.

Propose, implement and interpret a test of whether the effect of the subsidy on take up rates differed for household whose household head had monthly earnings above or below the mean monthly income prior to the treatment. Make sure to write down the regression equation with the interaction terms that you are using for your test.

In [ ]: