Program Evaluation Problem Set 2

Hye-Min Jung

5/14/2020

1.

- The most ideal experiment will be Randomized Controlled Trial. Because when RCT work, I can estimate treatment effect of FIONA on profits for the average farmer very easily.
- Necessary dataset from running RCT to answer the question:
  - data collected at the individual level
  - treatment status
  - pre-treatment observables characteristics
  - post-treatment observables characteristics
  - post-treatment outcome
- Define: $D_i \in \{0, 1\}$ as the treatment(taking rainfall-index insurance) indicator for farmer i
  - when farmer i is treated, $D_i = 1$
  - when farmer i is not treated, $D_i = 0$
- Define: $Y_i\left(D_i\right)$ as the profits as a function of D_i
  - when farmer i is treated, we observe profit $Y(1)$
  - when farmer i is not treated, we observe profit $Y(0)$
- Since under randomization, $\tau^{ATE} = E\left[Y_i|D_i = 1\right] - E\left[Y_i|D_i = 0\right]$, I can simply estimate the impact of treatment(taking rainfall-index insurance) for profit from data:
  - $\hat{\tau}^{ATE} = \overline{Y(1)} - \overline{Y(0)}$
  - To put it simply in words, difference in means between treated and control group in the data is the impact of the taking FIONA.
- Hence, 'observed effect of FIONA on profits, compared to the effect of FIONA on profits that we would have observed without the FIONA program for average farmer' can be estimated by RCT, the ideal experiment.

2.

- Instead, treatment effect for the treated units will allow us to answer the effect of FIONA on profits among farmers who 'took up insurance'.
- I can estimate $\tau^T = \frac{\bar{Y}(R=1) - \bar{Y}(R=0)}{P^{D_i=1}_{R_i=1} - P^{D_i=1}_{R_i=0}}$, using two treatment variables.
  - (1) Regress Y_i on R_i to recover $\hat{\tau}^{ITT}$

- (2) Regress D_i on R_i to recover $\hat{\pi}^C$
- (3) $\hat{\tau}^{LATE} = \frac{\hat{\tau}^{ITT}}{\hat{\pi}^C}$

- Unlike estimating average treatment effect for entire farms in (1), we are only looking into the treatment effect for the units that took up the insurance.
- We can see who is treated in the treatment group, but can't observe who would've been treated in the control group.

3.

- If we simply compare the post-treatment profit of FIONA to non-FIONA farms on average, we will be end up capturing treatment effect mixed up with different chraceteristics that is embeded in different districts.
  - (1) If there is selection on observables : FIONA farms were wealthier from the start
  - (2) If there is selection on unobservables : FIONA farms liked the insurance more than non-FIONA farms
  - (3) If there is selection on unobservables : FIONA farms had politicians that attracted goverment to offer the insurance.
- This can be problematic, because when we look at the estimate, we will not be able to distinguish between effects of insurance and the effect of different observable and unobservable characteristics on the outcome.

4.

- Observations for farmer_birth_year(1972, 1973) had string structure.
  - To fix this, these string structured observations (nineteen seventy-two, nineteen seventy-three) were recoded to integers(1972, 1973).
- **Potential approach 1: Regression adjustment**
  - To get $\hat{\tau} \approx \tau^{ATE}$, I will estimate $Y_i = \alpha + \tau D_i + \gamma X_i + \nu_i$
  - Regression adjustment design runs regression with selection on observables.
  - Applying this approach to FIONA, running regression profits on treatment plus couple of other covariates that has been observed allow us to estimate the impact of FIONA in profits.
- **Potential approach 2: Nearest-neighbor matching**
  - τ^{ATE} is estimated by calculating weighted average of $\bar{Y}_T - \bar{Y}_U$ for each cell $X = x$ that are uniquely defined by the covariates.
  - Applying this approach to FIONA, put treated and untreated farmers on observables together and compare them in terms of the profits by looking at the differences.
  - The profit differences between treated farmer and untreted farmer is treatment effect.
  - Farmer is a unit of analysis for FIONA application for both SOO designs.

5.

- Below, is the balance test table created to check the differences between FIONA and non-FIONA farmers on observable chracteristics, by running
  - $Y_i^{\text{baseline}} = \alpha + \tau D_i + \nu_i$

<Balance Test Table : Comparison of FIONA to non-FIONA farmers>

Variables	(1) FIONA farmers	(2) non-FIONA farmers	(1)-(2) Difference	P-value
2005 profits	20001.71	19989.878	11.835	0.491
birth year	1968.82	1968.93187	-0.11187	0.498
KARUR	0	0.200000	-0.200000	<2e-16
MADURAI	0	0.200000	-0.200000	<2e-16
PUDUKKOTTAI	0	0.200000	-0.200000	<2e-16
TENKASI	0	0.200000	-0.200000	<2e-16
THANJAVUR	1	-2.949e-15	1.000e+00	<2e-16
LENTILS	0.3	3.000e-01	1.415e-15	1
RICE	0.3788	0.400000	-0.021200	0.0604
WHEAT	0.3	3.000e-01	1.115e-16	1

– DINDIGUL, COTTON as baseline for district and crop respectively.

- FIONA and non-FIONA farmers turns out to be similar for observable characteristics (excluding for districts) given that the none of the coefficients for treatment variable were statistically significant in the balance test.
- However, FIONA and non-FIONA farmers in KKARUR, MADURAI, PUDUKKOTTAI, TENKASI, THANJAVUR were different from baseline district DINDIGUL.
  - This corresponds to our concern in (3). FIONA only impacted certain districts and non-FIONA districts were not offered any insurance products.
- Balance check result makes me feel better about the concern we had in (3).
  - Because once we control for this problematic distric covariates by adjusting regression, we can eliminate selection on observable (assuming that what we observe is all that matters).

6.

- Central assumption required for SOO design to be valid (Conditional independence)
  - $(Y_i(1), Y_i(0)) \perp D_i | X_i$
  - In put into words, potential outcomes are independent of D_i, conditional on covariates
  - This means, once we control for Xi, we've eliminated selection. Therefore, treatment is as good as random and we have "Strongly ignorable treatment assignment".
- Also, we need a second assumption (Common support),
  - $0 < \Pr(D_i = 1 | X_i = x) < 1$
  - To put into words, the probability that $D_i = 1$ for all levels of X_i is between 0 and 1.
  - This means, there are both treated and untreated units for each level of x.
- Under conditional independence and common support, we can get from τ^{SOO} to τ^{ATE}!
- Validity of common support assumptions:
  - Problem exists in farmer_birth_year. In some years there is no on being treated. But this is trivial for birth year as there are many more values for this covariate.
  - But the 'crop' covariate, all farmers harvest cotton received FIONA, while under other crops there are both FIONA and non-FIONA. So crops will be included selectively.
- Because common support assumption is broken, it is not likely that I will be obtaining credible estimate using the data.
  - Suppose we have cotton farmer and rice farmer in the population. This assumption is saying, that there are both treated cotton farmer and treated cotton farmer as well as untreated rice farmer and untreated rice farmer in the population.

– If that weren't true, imagine the population where the entire population treated were cotton farmer and our entire population of untreated were rice farmer, then there will be no way to separate treatment from crop factor, because there is perfect correlation between treatment and the type of crop grow.

7.

8.

9.