Problem Set 9

1. Derive the ATE decomposition result.

- $E[Y^1|D=1] - E[Y^0|D=0] = ATE + \ E[Y^0|D=1] - E[Y^0|D=0] + \big(1 - P(D=1)\big) * (ATT \ - ATU)$

2. Use Q2data.csv to solve this question

(a) Compute the ATE and report the value.

(b) Compute the ATT and report the value.

(c) Compute the ATU and report the value.

(d) Compute the difference between ATE and the estimate of $E[Y^1|D=1] - E[Y^0|D=0]$.

(e) Verify the decomposition result using this data.

(f) Give an example of a policy, whose ATE might not be equal to ATT.

3. Use nswre74.dta to solve this question. The dataset was provided by Rajeev Dehejia and it was used for his paper, "Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programs," Journal of the American Statistical Association, Vol. 94, No. 448 (December 1999), pp. 1053-1062.

   The data was downloaded from

   https://economics.mit.edu/faculty/angrist/data1/mhe/dehejia

The data includes sample who participated in the National Supported Work Demonstration (NSW) program, which is to support the workers without enough job skills to reenter the labor market. The program was randomized, so the treated group received the job training whereas the control group did not receive it. The variable "treat" indicates whether the sample is the treated group or not. The variables "re74" and "re75" are the pre-intervention earnings. The program was implemented between year 1975 and 1977. The variable "re78" is the post-intervention earnings.

(a) Run regressions to test whether the pre-intervention incomes are different between treatment and control group. Report either test statistic / standard error / p-value to support your conclusion.

(b) Test whether the treated and the control have different demographic characteristics (age, education, race, marital status). (Hint : use the "ttest" (or "estpost ttest") command. Look up the syntax by typing "help ttest"). Interpret the results and state any concern regarding identification.

(c) Run a regression to estimate the average treatment effect of the program using the post-intervention income variable, "re78". Explain why the regression estimand is the average treatment effect. Report either test statistic / standard error / p-value to support your conclusion.

(d) State the conclusion from the analysis. Was the program effective to raise the income of the workers?

4. State the assumptions to use the matching research design.

5. Use Pset2data.dta to solve this question. The data includes an outcome variable Y, treatment dummy D, and covariate X. Do matching using the covariate X.

   (a) Plot the histogram of X by the treatment dummy D. Comment on whether any assumption for matching is violated.

   (b) Report what fraction of observations can be exactly matched on X.

   (c) Do the exact matching on X and report the ATE with its standard error. Use the observations which can be exactly matched only.

(d) Do the nearest neighbor matching on X using 1 nearest neighbor. Use all observations in this calculation. Do not adjust for the bias. Report the ATE with its standard error.

(e) Repeat the (d) but with bias adjustment option. Report the ATE and its standard error.

(f) Do the Coarsened Exact Matching on X. Use the automatic binning done by "cem" command. Report what fraction of observations can be matched after binning.

(g) Estimate the ATT using the CEM weight. Report the standard error.
(NOTE : the STATA "cem" command gives a weight variable to compute the ATT. If you want to compute the ATE, use "cem_strata" variable with "teffects nnmatch" command.)

(h) Compute the propensity score using a logit model. Plot the histogram of the propensity score by treatment D.

(i) Use the Imbens-Hirano estimator to compute the ATE with the propensity score you computed in (h). Report the ATE with its standard error.

6. (Instrumental Variable) You will use data Pset3Q1.dta to solve this problem.

(i) Run an OLS regression to estimate the return to an additional year of schooling. Regress the weekly log wage onto educ and dummies of years of birth.

(ii) Instrument the years of schooling using the quarters of birth dummies. Report the return to an additional year of schooling with the standard error. Compare the estimate with the one obtained in (i) and explain why they are

different.

(iii) State whether your model is just identified or overidentified and explain why.

(iv) Test whether the instruments are weak using the Kleibergen-Paap F-statistic.

(v) Test whether you can reject the exogeneity of the instruments using the Hansen J-test.

(vi) Now, instrument the years of schooling using the quarters of birth X year dummies. Report the return to the years of schooling with the standard error.

(vii) Test whether the new instruments in (iv) are weak using the Kleibergen-Paap F-statistic.

(viii) Test whether you can reject the exogeneity of the new instruments in (iv) using the Hansen J-test.

7. (Fixed effect) You will use data Pset3Q2.dta to solve this problem.

   (i) Examine the data and state whether the data is a balanced or unbalanced panel.

   (ii) Run the following OLS regression and report the estimate of the marital premium for **males** with the standard error.

$$\log wage_{it} = \beta_0 + \beta_1 married_{it} + \beta_2 potential\ experience_{it} + u_{it}$$

   (iii) Explain why the OLS regression coefficient may be biased.

   (iv) Run the following fixed effect regression and report the estimate of the marital premium for **males** with the standard error. Cluster the standard error at the individual level.

$$\log wage_{it} = \beta_0 + \beta_1 married_{it} + \beta_2 potential\ experience_{it} + a_i + u_{it}$$

   (v) Explain why the FE regression may solve the concern in (iii).