Tutorial 1 Regression and matching to evaluate public policies

Exercise 1 : Estimating the effect of the National Supported Work using a randomized experiment (Lalonde, 1986 ; Dehejia and Wahba, 1999

The National Supported Work (NSW) is a labor market training program developed over the period 1975-1977 in the US. It aimed at providing work experience in different sectors (construction, restaurant, ...) for a period of 9-18 months to disadvantaged workers lacking basic job skills (ex-drug addicts, ex-criminal offenders, high-school dropouts). The NSW program assigned qualified applicants to training positions RANDOMLY. Those assigned to the treatment group benefited to the program while those assigned to the control group did not participate in the training program.

Since the cost per training participant is quite high (ranging from 6800$ to 9100$ in 1982 $, including the trainees' subsidized wages), the effect of the program on their future labour market trajectories have been evaluated. Both treated and non-treated individuals participated in follow-up interviews. The outcome variable of interest is post-intervention (1978) earnings.

The analysis focuses on the program effect on males participants' earnings (Dehejia and Wahba, 1999). The experimentally randomized data set includes 185 treated and 260 control observations.

a. Using the notations of the potential outcomes framework, write the expression of Average Treatment Effect on the Treated (ATT). What is the main issue to estimate this effect ?

b. Write a naive estimator to estimate the ATT. Is it an unbiased estimator ? Why ?

c. Write the corresponding OLS regression. Import the data set "nswre74.dta" with R and estimate the ATT.

d. Construct confidence intervals first assuming homoskedasticity and then relaxing this assumption.

e. Add other covariates in the regression (the race, the marital status and the number of years of education) and re-estimate the ATT. Is the coefficient expected to change after adjusting for these variables ? Why ?

f. What do you observe ?

Exercise 2 : Estimating the effect of the National Supported Work using a non-experimental sample (Lalonde, 1986 ; Dehejia and Wahba, 1999)

As Lalonde (1986) and Dehejia and Wahba (1999), use a non-experimental sample drawn from the Current Population Survey, including all males less than 55 years of age and check whether observational studies can replicate or not results obtained from randomized experiments.

a. Re-estimate the ATT using the data set "cps1re74" estimating a simple bivariate OLS regression. What do you observe ? Was it expected ? Why ?

b. Adjust for more covariates in the regression, controlling for the age, the age squared, the number of years of education, the race, the earnings in 1974, the earnings in 1975 and two dummies indicating whether the individual were working in 1974 or in 1975. Re-estimate the ATT ? What do you observe ?

c. Are distributions of covariates well balanced between the control and the treated group ?

d. Estimate the propensity score with a logit model using the age, the age squared, the cubic age, the number of years of education and its squared term, a dummy indicating whether the individual has at least one high-school degree, the marital status, the race, the earnings in 1974, in 1975, an interaction term between the earnings in 1974 and the number of years of education and two dummies indicating whether the individual were working in 1974 or in 1975 .

e. Plot the distribution of the propensity score in each group (control and treated). Comment.

f. To ensure overlap, apply the method suggested by Dehejia and Wahba (1999). How many control units have been discarded ?

g. To ensure overlap, apply the method suggested by Crump et al. (2009). How many control units have been discarded ? Look again at the balancing property. Comment.

Nearest-neighbour matching methods

h. From the sample obtained by the Dehejia and Wahba's method, apply a one-to-one propensity-score based nearest neighbour matching method with replacement and setting a caliper to 0.25 standard errors of the value of the propensity score to determine the ATT. Compare with your OLS estimate.

i. Compute the standard error of your ATT estimate by nearest neighbour matching by bootstrap.

j. What is the main issue of the propensity-score based matching estimators ? Which kind of matching estimator can be used to address this issue ? Use this estimator and compute the ATT. Compare with OLS estimates. Conclude.

Subclassification methods

k. From the sample obtained by the Dehejia and Wahba's method, apply a stratification estimator to compute the ATT. Compare with OLS estimates.

Weighting estimator

l. From the sample obtained by the Dehejia and Wahba's method, apply a Inverse Probability Weighting (IPW) estimator to compute the ATT. Compare with OLS estimates.