Econometrics 1

TD9

The goal is to assess the efficiency of the "Job Training Partnership Act" (JTPA). To do so, we follow the approach of Abadie, Angrist et Imbens in an article published in 2002, [1] and we use the data obtained from a randomized experiment implemented in the late 1980s. Henceforth, we are interested in the effect of the public policy on the sum of incomes over a duration of 30 months following the assignment to the treatment.

We use the same notation as in the course, and the variables `d`, `y`, and `z` in the dataset have the same meaning as in Chapter 5. In what follows, we assume $D(1) \geq D(0)$ (monotonicity) and $Z \perp\!\!\!\perp (Y(0), Y(1), D(0), D(1))$ (independence). Remember that a "complier" is defined as an individual such that $D(0) = 0, D(1) = 1$, and we denote this event C ; an "always taker" satisfies $D(0) = D(1) = 1$ (we denote this event AT) ; a "never taker" satisfies $D(0) = D(1) = 0$ (this event is denoted by NT).

1. Estimate the average effect of the treatment on the compliers : $\delta^C = E(Y(1) - Y(0)|C)$. Does the training given in the JTPA program seem effective ?
2. Compute the probabilities $P(C)$, $P(AT)$, and $P(NT)$ as a function of the distribution of observed variables. Is it possible, in some situations, to reject the hypotheses of monotonicity and independence ? Estimate the probabilities for the three groups in the data.
3. To better know who are the compliers, for some discrete random variable X, we study the quantities $g(x) = P(X = x \,|\, C)/P(X = x)$.
   Under the assumption that $Z \perp\!\!\!\perp (X, Y(0), Y(1), D(0), D(1))$, write $g(x)$ as a function of the distributions of the observed variables (D, Z, X).
   Estimate $g(x)$ in the following cases : X is the gender, X is the ethnic group, X is the level of education [2], and X is age categories. Comment on the results and describe the population of compliers.

1. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings", published in *Econometrica*. It can be found here : https ://economics.mit.edu/files/11871. The data used comes from the website of Joshua Angrist.

2. As regards the level of education, we use the variable `hsorged` in the dataset. It is equal to 1 if the individual has at least the equivalent of "baccalauréat" (achievement of high school), and 0 otherwise.

4. We define $\delta^{AT} := E(Y(1) - Y(0)|AT)$, $\delta^T := E(Y(1) - Y(0)|D = 1)$, and $\lambda := P(C, Z = 1)/[P(C, Z = 1) + P(AT)]$. Show the following equality :

$$\delta^T = \lambda\delta^C + (1 - \lambda)\delta^{AT}$$

   Estimate λ in the data. Explain why it is sensible here to assume that $(\delta^T - \delta^C)/\delta^C$ is close to 0.

5. In this question, we assume that $\delta^T = \delta^C$. Show that, under this assumption, we can retrieve the selection bias B relative to the "naive" difference : $E(Y|D = 1) - E(Y|D = 0)$. Estimate B with the data and comment.

6. Compute the estimators of the previous questions again (except those computed in Question 3) separately on the subsample of men and on the subsample of women. What do you notice ?

7. Henceforth, for Question 7, 8, and 9, we assume that $P(AT) = 0$. Show that the selection bias B satisfies :

$$B = \kappa[E(Y(0)|C) - E(Y(0)|NT)],$$

   where κ is a term to be determined.

8. Let assume that the individuals decide to be treated if the monetary gains $Y(1) - Y(0)$ exceed the cost c associated with the training. We assume that $(Y(0), Y(1) - Y(0) - c)$ is a Gaussian vector, that is, is distributed according to a bivariate Gaussian distribution. Express B as a function of κ, $\rho = \text{Corr}(Y(0), Y(1) - Y(0) - c)$, $\sigma_0 = V(Y(0))^{1/2}$, and $p = P(C)$.

9. How can you estimate ρ? Compute the estimate on the entire sample, then on the subsample of women and on the subsample of men separately. Discuss the results.