

STATA Exercise 1: A randomized experiment with intensive job search assistance.

February 23, 2018

Crépon et al. (2013)¹ consider a large-scale job search assistance program carried out 2007-2008 in France. The program targets young graduates with at least 2-year college who have been unemployed for at least 6 months. A two-step design is used for the experiment:

- In the first step, the proportion of job seekers to be assigned to the treatment is randomly drawn in each of the 235 labor markets (here: French cities participating in the experiment). The proportions are 0%, 25%, 50%, 75%, or 100%. To improve precision, Crépon et al. form groups of five agencies covered areas of similar size and with comparable populations. We refer to these 47 areas as quintuplets.
- In the second step, eligible job seekers in each labor market are randomly assigned to treatment following the proportion determined in the first step.

The treatment considered is carried out by private agencies that are contracted to provide intensive placement services. Job seekers assigned to treatment can decide whether or not to accept the offer of the intensive job search assistance by the private provider. If not assigned or if refusing the intensive job search assistance program, the job seeker will get the standard job search assistance by the counsellors of ANPE (French Public Employment Agency). The private provider is paid conditional on the individual finding a job with a contract for at least 6 months and staying employed for 6 months.

The data set contains 11,806 observations for a survey carried out 8 months after the randomization took place. All workers in the data set were regular unemployed or undeclared unemployed. The response rate is high, being 70.3% for non-assigned individuals and 72.2% for individuals who were assigned to treatment.

1. Calculate the average share of individuals with a long-term fixed contract (*LTFC*) for those assigned to treatment, \bar{y}_1 , and for those who were not

¹Crépon, B., E. Dufló, M. Gurgand, R. Rathelot, and P. Zamora (2013), "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment", *Quarterly Journal of Economics*, Vol. 128, No. 2, pp. 531-580.

assigned to treatment, \bar{Y}_0 . The treatment assignment variable is called *assigned*. It indicates whether individuals were assigned to treatment, but not whether they were actually treated. Compute the difference in means estimator $\bar{Y}_1 - \bar{Y}_0$ and use a two-sample t-test with unequal variances to test the difference in means. Repeat the analysis using long-term employment (*LT*).

2. How can we estimate the difference in means estimator with a linear regression?
3. Does $\bar{Y}_1 - \bar{Y}_0$ consistently estimate *ATE*? Does $\bar{Y}_1 - \bar{Y}_0$ consistently estimate *ATT*?
4. Do the results in questions 1-2 change when including controls? You should include dummies for the 47 quintuplets (the variable is also called *quintuplets*)², dummy for gender (*male*), past unemployment duration and its square (*pastudur* and *pastudursq*), 4 education dummies (*bac*, *bac3*, *bac4*, *bac5*)³ as well as 8 cohort dummies for *cohort3-cohort10*.
5. When estimating the effect of treatment assignment on employment, what condition should the control variables satisfy?
6. With a randomized assignment what is the reason to include explanatory variables?
7. We can also use the explanatory variables for balancing tests. These can be used as indicative evidence that the randomization works well and that small sample variations do not affect the estimates. Test whether the means of each of the explanatory variables (except the quintuplets dummies) differ between the unemployed assigned to treatment and the unemployed not assigned to treatment.
8. As an alternative (or complimentary) to the previous question, we can regress the assignment dummy on all the explanatory variables (including the quintuplets dummies), that is

$$Z_{ic} = X'_{ic}\alpha_1 + u_{ic}$$

where i indicates the individual and c the local labor market cluster. Z_{ic} is the assignment variable and X_{ic} is the explanatory variables specified above. Use a partial F-test to test for joint significance of the explanatory variables except the quintuplets dummies.

9. Recall that in some areas zero workers were assigned to treatment. We will include a dummy for being in an area with positive treatment share.

²You can include dummies by writing i.quintuplets.

³In France, the baccalauréat (or 'le bac') is an academic qualification which French students take at the end of high school. The numbers indicate the years of study on top of the baccalauréat. The final category *bac5* is for 5 or more years.

Denote this variable by W_{ic} (in the data set this variable is called *treatarea*) and consider the following model

$$Y_{ic} = \beta_0 + \beta_1 Z_{ic} + \beta_2 W_{ic} + X'_{ic} \gamma_2 + u_{ic} \quad (1)$$

As above i indicates the individual and c the local labor market cluster. Z_{ic} is the assignment variable and X_{ic} is the control variables specified above. Estimate the regression model and give interpretations of β_1 and β_2 and $\beta_1 + \beta_2$.

10. Examine if the results are similar for men and women.
11. Since only assignment to treatment is random, we are estimating the intention to treat (ITT) effect and not the treatment effect. Suppose you also have a dummy variable for being treated. Which type of estimation strategy [which you should know from a previous course] can be used to estimate the treatment effect?