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

February 23, 2018

1. Estimates of the difference in means estimator $\bar{d} = \bar{y}_1 - \bar{y}_0$

	$\bar{y}_0$	$\bar{y}_1$	$\bar{d}$	t-stat
LTFC	0.160	0.178	0.018	2.53
LT	0.365	0.382	0.017	1.89

   Hence, the intensive job search assistance by the private provider seems to increase both measures of employment, but whereas the increase in long-term fixed contract (LTFC) employment is significant at 1 percent, the increase in in long-term (LT) employment is only significant on a 10 percent level.

2. We can estimate identical results by, for example, regressing LTFC on the dummy variable for assignment to treatment (*assigned*), which we denote by Z_{ic}

$$y_{ic} = \beta_0 + \beta_1 Z_{ic} + u_{ic}$$

   The coefficient β_1 is the difference in means, whereas β_0 is the employment share for non-assigned individuals.

3. The answer is no. When we have randomization of treatment, we have $\bar{d}$ consistently estimates both $ATE = ATT$. We will show this below more formally, but we need to notice that being assigned to treatment is not the same as being given treatment. Some of the unemployed may refuse to take part in the intensive job search course. Therefore, we will only estimate the intention to treat effect (ITT).

   The difference in means estimator is given by $E(y|D=1) - E(y|D=0)$. When D is independent of (y_1, y_0), which will be the case when individuals are given treatment randomly, we can write

$$E(y|D=1) = E(y_0(1-D) + y_1 D|D=1) = E(y_1|D=1) = E(y_1)$$

   and

$$E(y|D=0) = E(y_0(1-D) + y_1 D|D=0) = E(y_0|D=0) = E(y_0)$$

so

$$E(y|D=1) - E(y|D=0) = E(y_1) - E(y_0) = ATE = ATT$$

Notice that when individuals are only assigned to treatment randomly and not given treatment randomly individuals can self–select themselves out of treatment. Since the intention to treat (ITT) uses the assignment variable, we also include those who were assigned, but did not participate in the program, in our estimate. If the effect of treatment has the same sign for all individuals, the effect of being assigned, but not participating should be numerically lower than the program effect. Hence, the intention treat estimator (ITT) will be closer to zero compared with ATE and ATT. Therefore, in cases where the treatment effect has the same sign for all individuals, it is sufficient to work with the ITT estimator as "the bias goes against you".

4. We are estimating the following equation

$$y_{ic} = \beta_0 + \beta_1 Z_{ic} + X_{ic}\gamma_2 + u_{ic}$$

where X_{ic} is specified in the question. We find the following estimates (we have excluded the estimates for the control variables)

	β_0	β_1
LTFC	0.101	0.017
	(2.06)	(2.49)
LT	0.308	0.015
	(4.88)	(1.70)

The estimates of the treatment effect (β_1) are very similar and are significant at, respectively, 5 and 10 percent (i.e. t-statistics are given in parentheses).

5. Let the assignment variable be denoted by Z. The explanatory variables should satisfy that conditional on X, Z and (y_1, y_0) are independent. In this case, where assignment is randomized, we would just need that the X's are mean independent of the error term. Usually, we should only include variables determined before treatment assignment (but notice that some of these variables may still be endogenous).

6. Crépon et al. (2013) include explanatory variables to show that small sample noise does not seem to influence the estimates.

7. It is important to notice that such types of test cannot be used to argue that treatment assignment is as good as random for an observational study. In case of perfect randomization, we should expect that 1 out of 20 of the differences to be statistically different from zero at a 5% level. However, since Crépon et al. only use 235 different randomized control areas, it is quite possible that, for example, the individuals in the areas treated, for

example, have higher levels of education. Based purely on these individual tests, a mechanical conclusion would be that the randomization does not work properly.

	$\bar{x}_{0j}$	$\bar{x}_{1j} - \bar{x}_{0j}$	$t - statistic$
pastudur	11.062	-0.154^{**}	-2.15
pastudursq	137.330	-3.634^{**}	-2.12
session3	0.132	0.001	0.23
session4	0.128	0.004	0.67
session5	0.110	-0.000	-0.05
session6	0.970	0.002	0.30
session7	0.126	0.004	0.70
session8	0.122	-0.003	-0.50
session9	0.081	-0.000	-0.09
session10	0.100	0.001	0.25
male	0.361	0.019^{**}	2.10
ba	0.052	-0.014^{***}	-3.63
ba3	0.182	-0.000	-0.03
ba4	0.109	0.018^{***}	3.08
ba5	0.183	-0.006	-0.89

8. As an alternative (or complimentary) to the previous question, we could regress the assignment dummy on all the explanatory variables (including the quintuplets dummies), that is

$$Z_{ic} = X_{ic}\alpha_1 + u_{ic}$$

We obtain a F-statistic of $F(15, 11744) = 2.62$ with an associated p-value of only 0.0006. Based on the answer to this and the previous questions, it seems that at least a long some dimensions, the randomization is not working properly and we have non-random selection into treatment. This essentially implies that we need to assume unconfoundedness and so the best thing to do is to include the explanatory variables. In this case, we are lucky as we just saw that this did not change the estimates.

9. Include a dummy for being in an area with a positive treatment share. Denote this variable by W_{ic} and consider the following model

$$y_{ic} = \beta_0 + \beta_1 Z_{ic} + \beta_2 W_{ic} + X_{ic}\gamma_2 + u_{ic} \tag{1}$$

Abstracting from the explanatory variables, the average share with long term fixed contract for the unemployed assigned to treatment will be $\beta_0 + \beta_1 + \beta_2$, whereas the average for individuals being in a treatment area, but not being treated is given by $\beta_0 + \beta_2$. Finally, average share being in a non-treated area is β_0. Therefore, β_1 gives the effect of being in a treatment area and being assigned compared to being in a treatment area and not being assigned. This effect is positive and significant at respectively 1% and 5%. $\beta_1 + \beta_2$ gives the net effect of program assignment.

Since this is insignificant, it appears that effect of the program seems to only redistribute jobs across unemployed in local labor markets. β_2 is the effect of being in a treatment area, so it is a measure of the (negative) externalities of the program. In other words, the stable unit treatment value assumption (SUTVA) does not seem to be met.

		LTFC	LT
β_1	Effect of being assigned	0.023	0.025
		(2.81)	(2.29)
β_2	Effect of being in a treatment area	−0.013	−0.021
		(−1.28)	(−1.56)
$\beta_1 + \beta_2$	Net effect of program assignment	0.010	0.003
		(1.06)	(0.28)

t-values in parentheses.

We can easily compute the estimate of $\beta_1 + \beta_2$. However, to obtain t-values, we reparameterize the model

$$\begin{aligned} y_{ic} &= \beta_0 + \beta_1 Z_{ic} + \beta_2 W_{ic} + X_{ic}\gamma_2 + u_{ic} \\ &= \beta_0 + \beta_1 Z_{ic} + \beta_2 Z_{ic} W_{ic} + \beta_2 (1 - Z_{ic}) W_{ic} + X_{ic}\gamma_2 + u_{ic} \\ &= \beta_0 + (\beta_1 + \beta_2) Z_{ic} + \beta_2 (1 - Z_{ic}) W_{ic} + X_{ic}\gamma_2 + u_{ic} \end{aligned}$$

where we have used that if an individual is assigned then she must be in a area with a positive treatment share, so $Z_{ic} W_{ic} = Z_{ic}$. Therefore, we estimate the following regression

$$y_{ic} = \alpha_0 + \alpha_1 Z_{ic} + \alpha_2 (1 - Z_{ic}) W_{ic} + X_{ic}\gamma_2 + u_{ic}$$

Then, $\alpha_0 = \beta_0$, $\alpha_1 = \beta_1 + \beta_2$ and $\alpha_2 = \beta_2$.

10. There are quite different results for men and women as effects for women are smaller and insignificant. This is not the common result in the literature, and in a Danish context counselling seems to work better for women than men. However, if looking at individuals with low levels of education, the effects are of similar size for men and women, but significant at 5% for men and only 10% for women (results are not given below, but can be found by running the STATA code). This partly reconciles the results with the literature, which not usually only consider individuals with a college degree.

	Men		Women	
	LTFC	LT	LTFC	LT
β_1	0.043	0.037	0.013	0.019
	(3.42)	(2.08)	(1.15)	(1.37)
β_2	−0.036	−0.043	−0.001	−0.010
	(−2.23)	(1.90)	(−0.05)	(−0.58)
$\beta_1 + \beta_2$	0.007	−0.006	0.012	0.009
	(0.50)	(−0.29)	(0.99)	(0.58)
Number of obs.	4387	4387	7419	7419

t-values in parentheses.

11. We could estimate the treatment effects using instrument variable methods. The first stage equation would be an indicator for treatment regressed against the assignment indicator (as well as the explanatory variables used in the second step). Since the assignment indicator should be random, we can use it as an instrument variable.