STATA Exercise 3: Regression Discontinuity Design and the Effect of Social Assistance Benefits on Employment. [ANSWERS]

April 9, 2019

1. From graphical inspection, it definitely looks like there is a discontinuity at 30 years of age in employment.

   However, we notice that there is quite some nonlinearity. Below, we will follow Lemieux and Milligan (2008) and just consider individuals aged 25 or more. This makes sense since then the estimate should not depend crucially on the type of specification.

2. When weighting the regression according to the cell number of observations, we obtain the identical estimates as running the regression on individual data. We estimate a 4.0 percentage points lower employment probability due to the higher social assistance benefit level. This effect is significant at a 1 percent level. This is a local average treatment effect, i.e. this estimate only holds for individuals around 30 years of age in Quebec. A similar increase in social assistance benefits could give different effect in another part of Canada (or the World) and for other ages. The LATE could be smaller or larger than ATE and ATT.

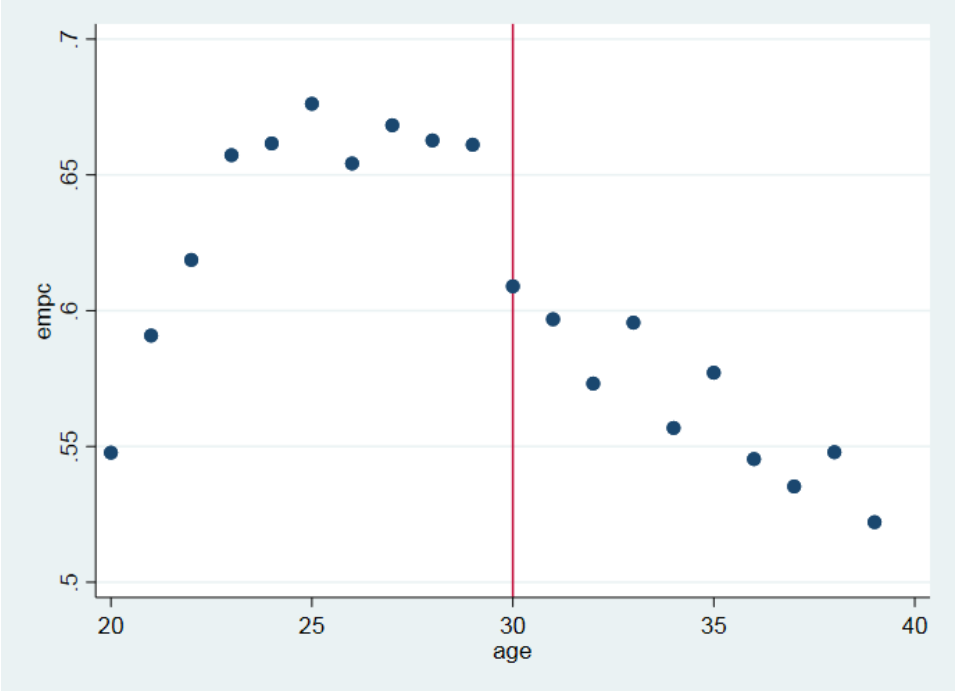
   We decide to use the usual OLS standard errors rather than robust standard errors since the former standard errors are largest. Thus, using the OLS standard errors seems to be the most conservative choice.

3. We obtain slightly larger estimates of -5.0% (squared) and -4.8% (cubic).

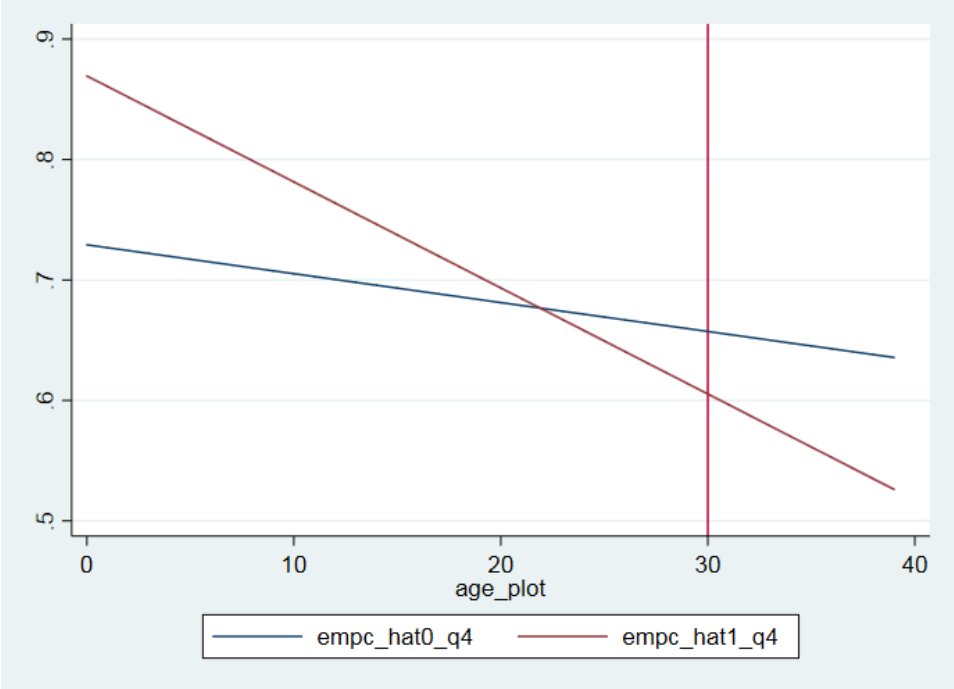
4. We also obtain slightly larger estimates using the donut RD: -4.7% (linear), -5.7% (quadratic), and -5.8% (cubic). This indicates that a significant share of individuals recorded with the age of 30 turn 30 years fairly late in the year and therefore receive the low social assistance benefit rate. The problem is, that we do not know the exact birthday of each individual.[1]

[1] Since we do not know the exact birthday, Lemeiux and Milligan (2008, p. 814) define the treatment indicator slightly different assuming a uniform distribution of birthdays and the day of the Census is June 1. This does not give much different results, so we will just use the simple treatment assignment throughout this exercise.

Question 1



Question 5



5. We construct two lines, one for $age < 30$ and for one $age \geq 30$. If we measure the difference between these two lines in $age = 30$, we get the right LATE estimate.

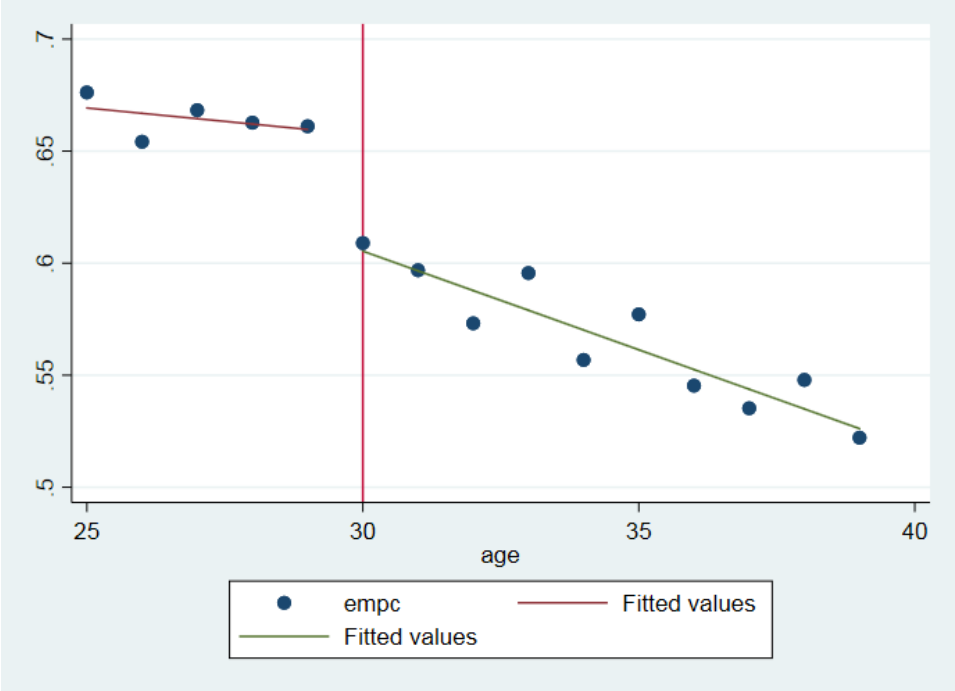
   However, what we accidently do here is that we measure the difference where $age = 0$. In this case, we have a fairly flat line to left of the discontinuity and to the right of the discontinuity a line with a quite negative slope. When extrapolating the lines to $age = 0$, we obtain that the line for the treated is way above the line for the non-treated. This way, the sign of the estimated effect changes and we obtain a meaningless estimate.

6. Centralizing the running variable at $age = 30$, we estimate an effect of similar size as before, -5.2 percentage points.

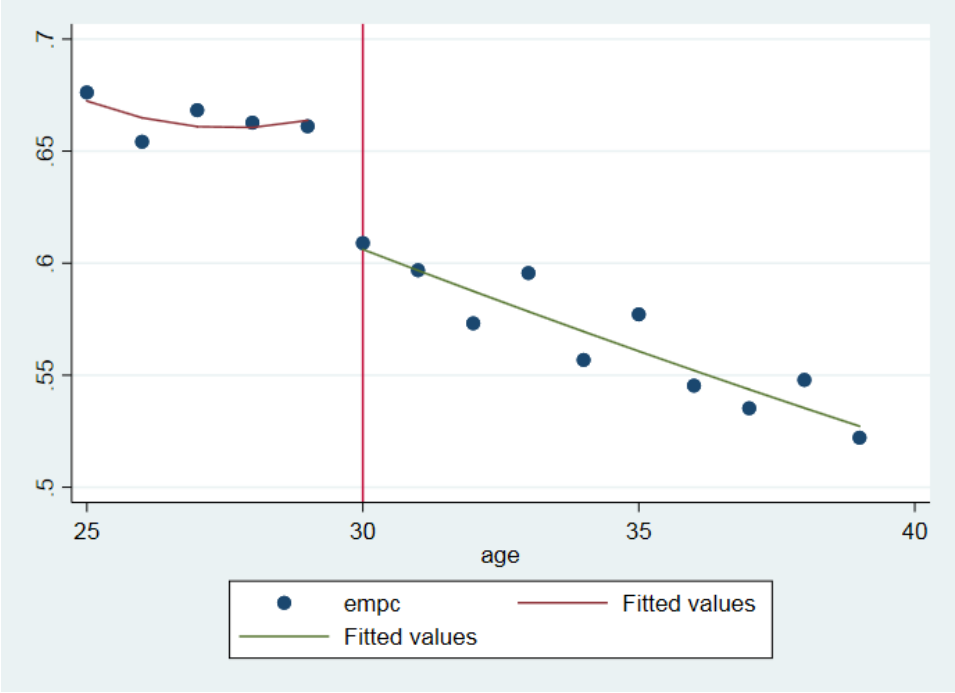
7. The graph illustrates why we obtain a slightly higher estimated effect: The slope to the left of the cutoff is fairly flat and previously we forced this slope to be the same as above 30, whereby we estimate a lower effect.

8. In both cases, we estimate negative effects, -6.4 and -2.2 percentage points, where the latter is when we also include observations for ages 20-24 years. The estimated effect of -2.2 percentage points is not significant even at a 10 percent level. Should we be concerned about this? No,

Question 7



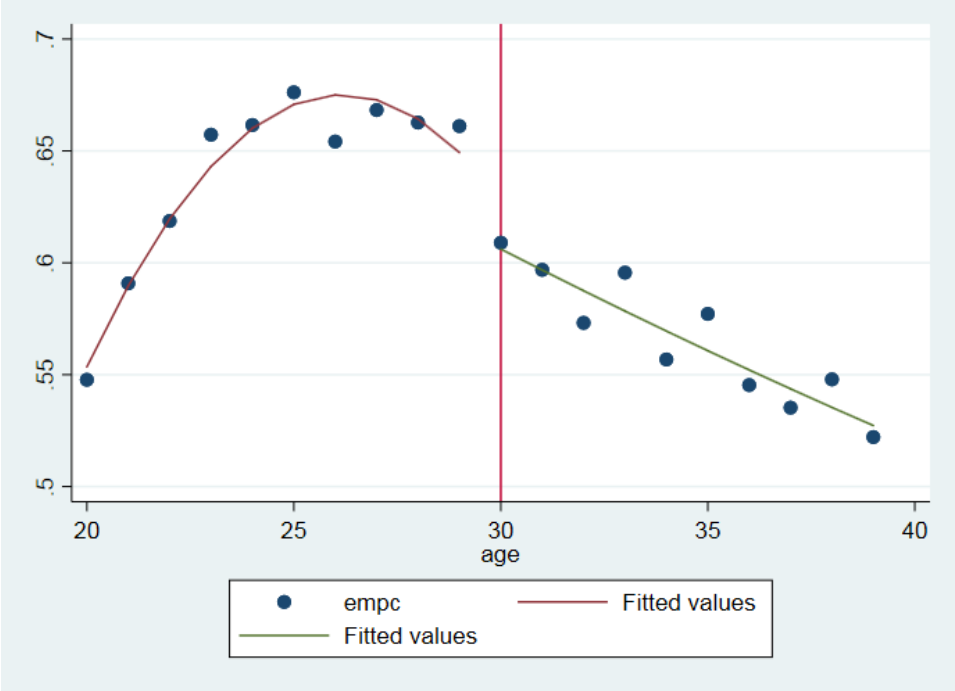
Question 8 i)



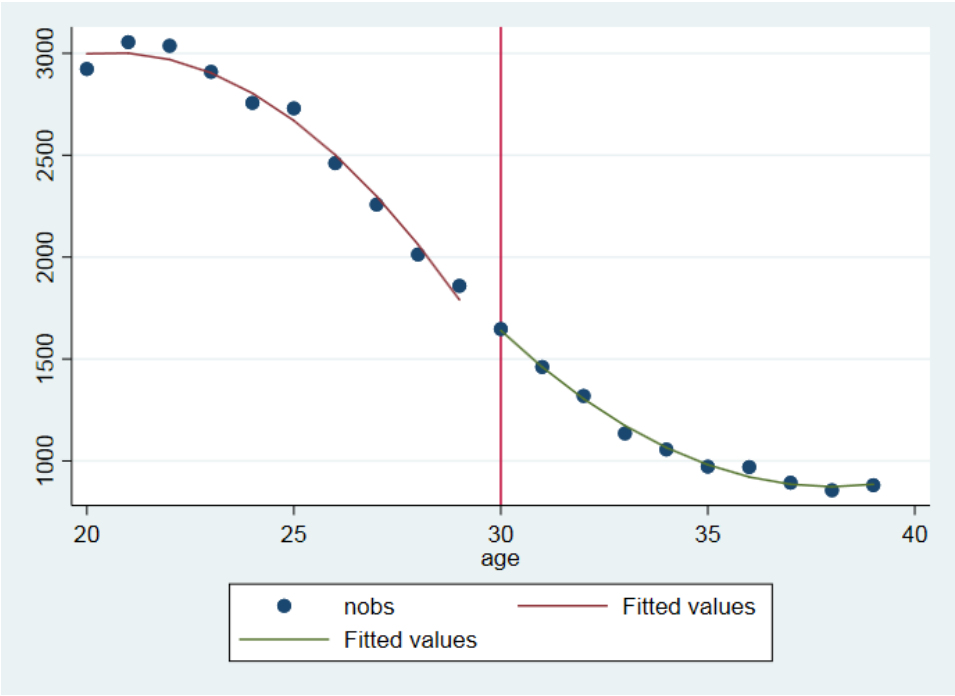
we are not concerned since the non-linearity, which reduced the estimate, is induced by including ages further away from the cutoff. This is not a concern since we are after a local estimate at the age of 30. It is important that we do not see a similar decrease in the estimate when allowing for squared terms when only considering persons aged $25 - 39$. Furthermore, it is important that the nonlinearities away from cutoff can be explained and does not visually look like the effect of the discontinuity.

9. We need to show that there is no bunching at the age of 30 years. Ideally, we would as Lemieux and Milligan (2008) use the share of individuals without children in Quebec and compare the share with the rest of Canada since it is possible that social assistance recipients in Quebec would decide to have children to avoid being paid the low level of benefits. We do not have this variable. Otherwise, it could also be nice to have the date of birth to be able to investigate the bunching more precisely. With the existing data, we just depict a scatter plot of the number of observations against the age. The graph suggests that there is no evidence of bunching at the age of 30 years. However, we obtain a significant effect at the cutoff in the regression used to obtain the fitted values. Luckily, this effect is insignificant (and even changes sign) using only observations for ages of 25 years and above. Hence, the significant effect on the full Quebec sample

Question 8 ii)



Question 9)



arises from the nonlinearity in the ages 20-24 and it is not a concern.

10. Only the effect at ages 30 and 31 are significant. However, the latter gives a lower effect (of -3.2 percentage points) and it is only significant at 10 percent.

11. Surprisingly, the estimated effect is significant at, respectively, 10 and 5 percent for Quebec in 1991 and for the rest of Canada in 1986. However, for Quebec in 1991 the effect is positive. Finally, for the rest of Canada in 1991, the estimated effect is positive and insignificant. It is a bit worrying to obtain the negative effect significant at 5 percent for the rest of Canada in 1986, but notice that the estimate is much lower than the estimates for Quebec in 1986.

12. We can use three different diff-in-diff strategies by exploiting different variation in the data. First, we estimate the following model on individuals aged $20-29$ years

$$Empc = \beta_0 + \beta_1 d_{\{quebec=1\}} + \beta_2 d_{\{year=1986\}} + \beta_3 d_{\{quebec=1\}} d_{\{year=1986\}} + u \quad (1)$$

where β_3 is the treatment effect. Notice that here we define the treatment as having a lower benefit level whereas we in the regression discontinuity

analysis defined the treatment as having a higher benefit level. Furthermore, notice that the effect we estimate is a ATT for all individuals aged 20-29.[2] In this case the identifying assumption is that the macroeconomic developments between 1986 and 1991 affected Quebec and the rest of Canada in a parallel way. We would need more time periods to show whether the assumption of parallel trends is likely to be met for Quebec vs. the rest of Canada. We obtain an insignificant estimate of -1.6 percentage points, that is the opposite sign as we expected. This is likely to be due to the parallel trends assumption is being violated. In other words, it was probably the case that the 1991 recession affected Quebec more than the rest of Canada. Hence, we suspect that the rest of Canada is a poor control group in this case. We can check whether Quebec was affected harder by the recession in 1991 by running equation (1) on individuals aged $30-39$ years. If we have parallel trends, we should expect that β_3 is zero. However, we estimate β_3 to -5.4 (significant at a 1 percent level), so we conclude that the parallel trends assumption seems to be violated.

Second, we can estimate the following model on data from 1986

$$Empc = \beta_0 + \beta_1 d_{\{quebec=1\}} + \beta_2 d_{\{age<30\}} + \beta_3 d_{\{quebec=1\}} d_{\{age<30\}} + u$$

Again β_3 is the treatment effect. Here, the identifying assumption is that the differences in the labor market conditions between individuals aged below and above age 30 are similar in Quebec and the rest of Canada. In this case, we estimate a significant effects of 6.8 percentage points, which is significant at 5 percent.

Third, we can estimate the following model on data from Quebec only

$$Empc = \beta_0 + \beta_1 d_{\{year=1986\}} + \beta_2 d_{\{age<30\}} + \beta_3 d_{\{year=1986\}} d_{\{age<30\}} + u$$

Again β_3 is the treatment effect. Here, the identifying assumption is that the local labor market conditions in Quebec were affected similarly for persons aged below and above 30 years between 1986 and 1991. We obtain an estimate of 6.9 percentage points, which is significant at 1 percent.

[2] Lemiuex and Milligan (2008) restrict attention to ages 29 and 30 to relate the regression discontinuity design to the diff-in-diff model.