PC session 3

RDD and DiD estimation

Home assignment: 1, 2, 3.
Due date: prior to 10am (Swiss time) on 19 May 2016.
Please email your documented solutions, including calculation/regression results, in a pdf-file to: martin.huber@unifr.ch. Group papers of maximum 4 students are permitted.
If possible, please bring your own laptops for the PC sessions.

Sharp and fuzzy RDD

McCrary and Royer (2011) investigate, inter alia, the effect of school entry policies on female education. They use natality data for Texas and California coming from each state’s Department of Health. At birth, each mother along with her health care provider completes an extensive survey about maternal and paternal demographic characteristics, maternal behaviors during pregnancy (e.g. prenatal care), and the health of the infant at birth.

The dataset RDDbirthday.RData includes 20000 observations for the year 1989. Your parameter of interest is the effect of being born prior to or after the school entry date (determining whether one enters school in the same year when turning 6 years old or one year later) on the outcome "mother’s years of schooling" (variable "mom_educ"). The variable “post” is a dummy for being born prior to or after the school entry date. Note that the running variable "R" provides the number of days born prior to or after the school entry date (it is negative if born prior and positive if born after).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>Number of days mother was born prior to/after the school entry date</td>
</tr>
<tr>
<td>mom_educ</td>
<td>Mother’s years of schooling</td>
</tr>
<tr>
<td>black</td>
<td>Dummy for mother being black</td>
</tr>
<tr>
<td>numvisits</td>
<td>Number of doctor visits prior to child birth</td>
</tr>
<tr>
<td>post</td>
<td>Dummy for being born prior to or after the school entry date.</td>
</tr>
</tbody>
</table>

1) Run a linear RDD estimation to evaluate the effect of being born prior to the age cut-off on mother's years of schooling. Use the total sample and assume that the regression function of the outcome (“mom_educ”) given the running variable “R” has the same slope on both sides of the cut-off defined by the school entry date. Comment on the results.

2) Run a linear RDD estimation using the total sample, with possibly different slopes of the regression function of “mom_educ” given “R” on both sides of the cut-off. Comment on the results.

3) Run a quadratic RDD estimation using the total sample, with possibly different slopes of the regression function of “mom_educ” given “R” on both sides of the cut-off. Comment on the results.
4) Run RDD estimation based on a local linear regression and a cross-validated bandwidth using a uniform kernel function. Use the “rdbwselect_2014” and “rdrobust” commands.

5) Now use a triangular kernel which gives less weight to observations further away from the cut-off. How does this affect the estimate?

6) Use the cross-validated bandwidth for a local linear regression with a triangular kernel to run RDD estimation by means of the “RDestimate” command in the rdd package.

7) Run RDD estimation including "black" as an additional regressor using the “RDestimate” command.

8) Use a quadratic regression model of the outcome “mom_edue” in the running variable “R” without controlling for other mother’s characteristics, using the “rdbwselect_2014” and “rdrobust” commands.

9) Test for discontinuities at the threshold with default values for the binwidth and bandwidth. (Hint: “DCdensity” returns the p-value)

10) Check further statistics such as the estimated log difference at the threshold (theta), standard error (se), binsize, and bandwidth (bw) by setting ext.out=TRUE.

11) Test for discontinuities at the threshold using a larger binsize and bandwidth than obtained in (10). What do you conclude when comparing the results to (9)?

12) Create a treatment variable “higheduc” which takes the value one if mother’s education is more than 9 years and zero otherwise.

13) Run fuzzy RDD estimation to evaluate the effect of “higheduc” on “numvisits” using a local linear regression with a triangular kernel by means of the “rdbwselect_2014” and “rdrobust” commands.

14) Use the cross-validated bandwidth for a local linear regression with a triangular kernel in the “RDestimate” command to perform fuzzy RDD estimation.

15) Run fuzzy RDD estimation when including “black” as an additional regressor.

**Difference-in-Difference**

Montalvo (2011) analyzes the electoral impact of the terrorist attacks of 11 March 2004 in Madrid. The data include region-level observations for congressional elections between 1989 and 2004 for the treated group (Spanish residents) and the control group (Spanish citizens in a foreign country).

Use dataset votingbombing.RData with the following variables:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>year</td>
<td>Year 1989-2004</td>
</tr>
<tr>
<td>dtreat</td>
<td>Dummy variable for treated which takes the value one for Spanish residents and zero for Spanish citizens abroad</td>
</tr>
<tr>
<td>pp</td>
<td>Percentage of votes for conservatives</td>
</tr>
<tr>
<td>psoe</td>
<td>Percentage of votes for social democrats</td>
</tr>
<tr>
<td>voters</td>
<td>Number of voters in the region</td>
</tr>
<tr>
<td>id</td>
<td>Province code</td>
</tr>
</tbody>
</table>
16) Perform a DiD estimation to evaluate the effect of the Madrid bombings on the ratio of votes for conservatives vs social democrats controlling for the number of voters in the region. (Hint: 2000 is the last pre-treatment period, 2004 is the post-treatment period). Report heteroskedasticity robust standard errors.

17) Estimate clustered standard errors (clustered on the regional level) using multiwayvcov package in R. How does clustering affect inference?

18) Conduct a placebo test by comparing Spanish residents and Spanish citizens abroad in two pre-treatment periods. Use conventional and clustered standard errors. Would you expect the coefficient on the interaction term to be significant? Why (not)?

19) Use DiD based on inverse probability weighting on propensity score to estimate the effect of the Madrid bombings on the ratio of votes for conservatives vs social democrats when controlling for the number of voters. Use bootstrapped standard errors with and without clustering.

**Conceptual questions**

20) Describe the DiD method. Which is the main identifying assumption and how can it solve the problem of treatment endogeneity?

21) In which case can covariates help to identify the causal effect with DiD?

22) Provide the intuition of the sharp RDD.

23) What is the key identifying assumption behind the fuzzy RDD?