UNIVERSITY OF MANNHEIM

E 5040 Impact Evaluation – Tutorial 6
Mathilda Featherston-Lardeux
mathilda.featherston-lardeux@uni-mannheim.de

Computer Session 6 – RDD and Non-parametric regression

Part 1: RDD

In this section, we try to estimate the effect of extended benefits for unemployed workers on the duration of their unemployment spell, based on Lalive (2008). The dataset "Tutorial6_data1" is a subsample of the true dataset used in this paper and contains the duration of unemployment of Austrian workers. An extension of the benefits was introduced in Austria, but only people of 50 years and older could benefit from it.

1. Why a RDD may be an appropriate design here? Is it a sharp or a fuzzy design? Are other impact evaluation designs seen in class also appropriate?

2. Reminder: Functional forms of the following kernels

   Epanechnikov: $K[z] = \begin{cases} \frac{\frac{3}{4}\left(1-\frac{1}{5}z^2\right)}{\sqrt{5}} & if\ |z| < \sqrt{5} \\ 0 & otherwise \end{cases}$

   Triangular: $: K[z] = \begin{cases} 1 - |z| & if\ |z| < 1 \\ 0 & otherwise \end{cases}$

Estimate the kernel weight using alternatively triangular, normal and epanechnikov kernels, with a bandwidth equal to 2.

3. Plot the kernel weights against age. What do you observe?

4. Create a variable for $z_i - z_0$.

5. Estimate y^+ and y^- using the local linear regression and triangular kernel weights.

6. Plot the expected value of the two regressions against age. Do they behave as expected?

7. Estimate the treatment effect using the RD estimator.

8. Interpret the result. What is the impact of the program for the overall population of unemployed?

9. Reproduce the results of questions 5 and 6 using the "rd" package.

10. Compute the optimal bandwidth.

11. Estimate the effects with different bandwidths and plot these estimates versus bandwidths.

12. With the available variables, what kind of placebo test could be performed to enhance the credibility of these results?

13. Suggest two other tests that could be performed if more variables were available.

Mathilda Featherston-Lardeux
mathilda.featherston-lardeux@uni-mannheim.de

Part 1: Non-parametric Regression

This section will be based on the Prestige of Canadian Occupations database. The observations are occupations. The dataset is "Tutorial6_data2".
Source: Canada (1971) Census of Canada. Vol. 3, Part 6. Statistics Canada [pp. 19-1–19-21]

A- Kernel Density Estimation

1- On the same graph, plot an histogram displaying the density of the variable "prestige" and its kernel density.
2- Estimate the kernel density of "prestige" using a bandwidth of 1.
3- Estimate the kernel density of "prestige" using a bandwidth of 20.
4- Estimate the kernel density of "prestige" using a Gaussian kernel function.
5- Estimate the kernel density of "prestige" using a rectangle kernel function.

B- Kernel regressions

1- Plot a scatter plot of "prestige" versus "income". What is the form of the relationship between the two variables suggested by this graph?
2- Regress "prestige" on "income" using OLS and add a line of predicted values to the previous scatter plot. Comment on this graph.
3- Using the command "npregress", estimate the non-parametric regression of "prestige" on "income" using kernel, bandwidths of 2000 and bootstrapped standards errors with 100 replications.
4- Plot the scatter plot of "prestige" versus "income" and the predicted line, using the command "npgraph". Save the graph in memory.
5- Re-estimate the last regression using bandwidths of 5000. Plot the result and save the graph in memory.
6- Re-estimate the last regression without specifying the bandwidths. How are they calculated? Plot the result and save the graph in memory.
7- Re-estimate the last regression adding the option "imaic". What is this option useful for? Plot the result and save the graph in memory.
8- Plot the four last graphs on one frame and comment.
9- Estimate the non-parametric regression of "prestige" on "income", "women" and "education" using kernel, bandwidths determined by cross-validation and bootstrapped standards errors with 20 replications.

10- Estimate the marginal effects of "women" on "prestige" for different values of "women". (hint: use the "contrast" option in "margins"). Plot the result.

11- Estimate the non-parametric regression of occupational prestige on income and the type of occupation using kernel, bandwidths determined by cross-validation and bootstrapped standards errors with 20 replications.

12- Estimate the marginal effects of the type of occupation on occupational prestige. Interpret the results. Plot the result.

13- Estimate the expected mean of occupational prestige for different values of income, by type of occupation. Plot the result.

C- Local polynomial regression

14- Using the command "lpoly", estimate a local average non-parametric regression of "prestige" on "income", displaying the confidence bands. Save the graph in memory.

15- Estimate a local polynomial non-parametric regression of "prestige" on "income" of degree three. Display the confidence bands and save the graph in memory.

16- Re-estimate the last regression specifying a bandwidth of 2000. Display the confidence bands and save the graph in memory.

17- Plot the last three graphs, in addition to graphs 3 and 4, in the same frame. Comment.

D- Locally weighted regression

18- Using the command "lowess", estimate a locally weighted nonparametric regression of "prestige" on "income". Save the graph in memory.