

Problem Set 6: Regression Discontinuity

Claire Duquennois

Name:

1 Empirical Analysis using Data from Manacorda, Miguel, & Vigorito (2011, American Economic Journal: Applied Economics)

This exercise uses data from Manacorda, Miguel, & Vigorito's paper, "Government Transfers and Political Support," published in the *American Economic Journal: Applied Economics* in 2011. This paper studies how receipt of a government anti-poverty cash transfer changes how beneficiary households support and view the government.

2 Finding the data

The data can be found on Edward Miguel's faculty website. Download and extract the contents from the `Government_Transfers_replication.zip` file.

3 Set up and constructing the data

The original data used in the paper is confidential. The authors instead provide the `reg_panes.dta` data file which is anonymized and created from the original data.

3.1 Question: Loading the Packages

Load any R packages you will be using: **Code:**

3.2 Question: Open the `reg_panes.dta` file. To complete this problem set you will need the following variables from this data file:

Name	Description
<code>aprobado</code>	Ever received PANES 2005-2007
<code>untracked07</code>	Untracked in 2007
<code>h_89</code>	Supports current government 2007 [1 to 3]
<code>hv34</code>	Supports current government 2008 [1 to 3]
<code>ind_reest</code>	Predicted Income
<code>newtreat</code>	PANES eligibility
<code>geo</code>	Geographic locality
<code>bl_medad</code>	Mean age
<code>lnbl_ytoth_pc</code>	Log per capita income
<code>bl_hhsize</code>	Mean household size
<code>bl_meduc</code>	Mean education
<code>missbl_medad</code>	Missing mean age
<code>misslnbl_ytoth_pc</code>	Missing log per capita income
<code>missbl_hhsize</code>	Missing mean household size
<code>missbl_meduc</code>	Missing mean education
<code>sexo</code>	Respondent is female
<code>edad</code>	Respondent age
<code>aniosed07</code>	Education in 2007
<code>misssexo</code>	Missing gender
<code>missedad</code>	Missing age
<code>missaniosed</code>	Missing education

Drop all other variables. If needed, give the variables you are keeping more intuitive names.

Code:

3.3 Question: The data as downloaded will require that you clean the variables of interest and construct a new dataset to generate the graphs. Start by generating the following cleaned variable:

-An indicator for receiving PANES that is NA if a respondent is untracked in 2007

Code:

3.4 Question: We are going to re-scale the variables that indicate support for the current government so that responses range from 0 to 1. To do this, tabulate the current variable to see how it is distributed and then generate a variable that will be NA if it is currently coded as 9, 0 if currently 2, 0.5 if currently 1 and 1 if currently 3. Do this for both the 2007 and 2008 variable.

Note: This is how the authors modify this variable in their code. It seems counter intuitive and does not correspond to the description of how this variable is coded in the survey questionnaire as reported in their appendix though it does correspond to their discussion in footnote 12. My guess is the transcription/translation of the survey question is incorrect.

Code:

3.5 Question: Generate a variable that is the square of predicted income.

Code:

- 4 We start by reproducing the main figures (2,3,and 4) of the paper as good figures are key to any regression discontinuity paper.
- 4.1 Question: The data consists of over 3000 observations. How many points are plotted on these figures? How should we interpret the y axis? What does each point below the threshold represent? What does each point above the threshold represent?

Answer:

4.2 Question: Why is the number of points above the threshold different from the number below?

Answer:

4.3 Question: Replicating these figures will require restructuring our data and calculating the values that are plotted. Generate a variable that will indicate the percentile group the observation is in. Note the difference in the number of percentile groups above and below the threshold.

Note: you may find the `xtile` function in R useful.

Code:

4.4 Question: For each of the percentile groups, calculate the mean of each of the variables we will use for plotting: predicted income, receipt of PANES, support for the government in 2007, and support for the government in 2008.

Code:

4.5 Question: Replicate figure 2. Make the figure as clear and informative as possible. You may want to create an indicator variable for percentiles above and below the threshold.

Code:

4.6 Question: What is the purpose of this figure and what should we take away from it?

Answer:

4.7 Question: (2 pages) Replicate figures 3 and 4. Make these figures as clear and informative as possible.

Code:

4.8 Question: Interpret these figures. What should we take away from them?

Answer:

4.9 Question: Replicate the results of the three regressions estimated in the first column of table 1. Present your results in a table. Interpret the coefficients.

Code:

Answer:

4.10 Question: Write down the specifications used in row 2 of columns 1,2 and 3 of table 1.

Answer:

4.11 Question: (2 pages) Replicate all of the results reported in row 2 of Table 1. Explain the difference between these specifications and interpret their coefficients.

Hint: the variables listed in the table above after newtreat are the controls you will want to include.

Code:

Answer:

4.12 Question: What is the point of including all of these specifications?

Answer:

4.13 Question: Using the coefficients estimated above, write out the function you would use to predict the probability a household supports the current government based on their predicted income score:

- a) If they are eligible for the transfer using the results from column 1.
- b) If they are not eligible for the transfer using the results from column 1.
- c) If they are eligible for the transfer using the results from column 2.
- d) If they are not eligible for the transfer using the results from column 2.
- e) If they are eligible for the transfer using the results from column 3.
- f) If they are not eligible for the transfer using the results from column 3.

Answer:

4.14 Question: How narrow is the “bandwidth” used by the authors. Why does this matter? Check that the results are robust to a narrower bandwidth.

Answer:

4.15 Question: The authors attribute these effects to the causal effect of receiving the government transfers. What is the implied assumption behind this interpretation?

Answer:

4.16 Question: What evidence do they provide to support this assumption?

Answer:

4.17 Question: Was this threshold eligibility score specifically designed for this particular program? Why does this matter?

Answer:

5 Submission instructions:

- Make sure the final version of your assignment is knit in pdf format and uploaded to gradescope. Make sure you have one question response per page (unless otherwise indicated) so that question positions align with the template in gradescope. The final PDF should be 25 pages long.