

Policy Evaluation Using Regression Discontinuity Designs: Exercise Sheet

Alexandre Grellet

Introduction

This exercise sheet is based on the paper by Meyersson (2014). It is designed to guide you from understanding the research context and methodology (Regression Discontinuity Design) to applying these methods in a replication exercise.

You are expected to blend intuitive explanations with technical details. The exercise is divided into four parts: understanding the research context, reviewing the methodology, replicating the results, and critically evaluating the findings.

Understanding the Research Context and Main Findings

Research Question & Institutional Context

Q1: Summarize in your own words the primary research question addressed in the paper. What concerns do the authors raise about the impact of Islamic political control on women's empowerment?

Q2: Describe the institutional context in Turkey as presented in the paper. Why is Turkey considered a suitable setting for studying the effects of Islamic political rule on educational outcomes?

Outcomes and Raw Associations

Q3: Identify the main outcome variables examined by Meyersson. Which educational measures are used (e.g., high school completion, enrollment), and for which age cohorts?

Q4: The paper contrasts raw (unconditional) correlations with causal estimates obtained via the RD design. What is the observed sign and magnitude of these raw associations, and how do they change when an RD design is implemented?

Part II. Methodology: Regression Discontinuity Design (RDD)

RD Design and Identification Strategy

Q5: Explain the fundamental idea behind a regression discontinuity design (RDD). What is the “forcing variable” in this study, and how is the cutoff defined?

Q6: The baseline specification in the model is the following:

$$y_i = \alpha + \beta m_i + f(x_i) + \varepsilon_i$$

- What does y_i represent?
- How is the treatment indicator m_i determined?
- What role does the function $f(x_i)$ play in the estimation?

Q7: In the present case, is the RDD fuzzy or sharp?

Local Linear Regression

Q7: Discuss why a local linear regression (and the choice of bandwidth) is crucial in RDD applications. *Note: in the paper, Meyersson uses the Imbens–Kalyanaraman (IK) optimal bandwidth selector from Imbens and Kalyanaraman (2012) , but later in the code we will just play with the bandwidth manually.*

Part III. Practical Replication of the Main Results

Replication Setup

```
# Load required packages
library(haven)
library(dplyr)
library(ggplot2)
library(rdrobust)
```

```
library(rddensity)
```

```
df <- read_dta("Meyersson.dta")
```

The following variables will be important

- Y: The primary outcome.
- X: The running variable (Islamic margin of victory).
- T: The treatment indicator.
- Other variables (e.g., hischshr1520m, vshr_islam1994) represent educational, political, and demographic characteristics.

Descriptive statistics

Q12: Propose a summary table of the following characteristics:

```
vars <- c("Y", "X", "T", "hischshr1520m", "vshr_islam1994",  
         "partycount", "lpop1994", "ageshr19", "ageshr60",  
         "sexr", "shhs", "merkezi", "merkezp", "subbuyuk")
```

Plotting the data around the threshold

Before we dig into the analysis, we should plot the information to have a visual idea about whether a regression discontinuity design could work.

Q13: Do a scatterplot of Islamic margins against the outcome variable.

Q14: Do raw comparison of means (polynomial degree 0) below and above the threshold. Is there any difference? *Hint: A useful package is `rdrrobust`*

```
library(rdrobust)  
rdplot(y = df$`outcome`, x = df$`forcing variable`, c = threshold, p = polynomial degree,  
       x.label = "label",  
       y.label = "label")
```

Q14: Try to graph the relationship with higher-order polynomials (1,2,3)

Q15: Now restrict the analysis to observations with $|X| \leq 50$.

Running the estimation

Q16: Restrict the analysis to observations with $|X| \leq 20$. Run two separate regressions (without controls) on the right and on the left of the threshold. Then give the RDD estimate.

Q17: Run a regression with an interaction term to estimate the RDD.

Q18: Add covariates to the regression.

Robustness checks (optional)

Q19: Estimate RDD using as an outcome a variable for which there should be no effect.

Q20: Perform a misspecification check using the function `rddensity`.

References

- Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3): 933–59. <https://doi.org/10.1093/restud/rdr043>.
- Meyersson, Erik. 2014. "Islamic Rule and the Empowerment of the Poor and Pious." *Econometrica* 82 (1): 229–69. <https://doi.org/10.3982/ECTA9878>.