

# READ ME FOR REPLICATION FILES

## Bounds on Treatment Effects in Regression Discontinuity Designs with a Manipulated Running Variable

François Gerard, Miikka Rokkanen, and Christoph Rothe

The Replication.zip file contains the programs used to build the dataset and generate the results in the paper. It also contains the results themselves. This document describes the steps necessary for replication and data access.

### 1. Directory structure

The construction of the dataset used in the empirical analysis was done using Stata 15. All the codes are included in the subfolder “Codes for data preparation” in the folder “Stata Analysis.”

The graphs in Figure 2 in the paper, as well as those in Figure 4, Figure 5, and Figure 6 in the Web Appendix were created using Stata 15. The code creating these graphs (“Create\_graphs.do”) and the log file from running that code (“Create\_graphs.smcl”) are included in the subfolder “Codes for results” in the folder “Stata Analysis.” The results generated by running that code is included in the subfolder “Graphs” in the folder “Stata Analysis”.

The results in Table 2 and Figure 3 in the paper, as well as those in Table 3 in the Web Appendix were created using R version 3.5.1. The code creating these results (“run\_brazil\_qe.R”), the R package implementing the methods developed in the paper (“rdboundsv1.01.R”), a copy-paste of the R console output from running that code (“run\_brazil\_qe\_log.txt”) are included in the subfolder “Codes for results” in the folder “R analysis.” That subfolder also includes a Stata code (“Create\_figure3.do”) creating the graphs in Figure 3 using the output from running the R code (“Create\_figure3.smcl” is the log file from running that Stata code). All the results generated from running these codes are included in the subfolder “results” in the folder “R analysis”.

Note that the latest version of the R and Stata packages implementing our methods can always be found here: <https://github.com/francoisgerard/rdbounds/>.

### 2. Stata analysis/Codes for data preparation

As described in the paper, the data used in the empirical analysis comes from two administrative datasets. The first one is a longitudinal matched employee-employer dataset covering by law the universe of formal employees. The second dataset is the registry of all UI payments. These datasets were obtained from the Brazilian Ministry of Labor and are proprietary. As a result, we are not able to make them available to researchers at large. If anyone would like to replicate the

results, there are two pathways. One is to contact the authors to seek an authorization to come on-site in order to use a secure server for the purpose of research integrity and verification only. An alternative is to seek data access through a data-sharing agreement with the relevant Brazilian administration. Other scholars have gained access to similar data for research purpose in the past. For this second option, we include all the codes that we used for the data preparation.

- A. A first set of codes prepares the data from the matched employee-employer dataset and the UI registry for the analysis of displaced formal workers' behaviors in relation to UI and formal reemployment outcomes (they were not created specifically for this paper)
  - "Data\_prep\_A1:" selects displaced workers in matched employee-employer data
  - "Data\_prep\_A2:" organizes displaced workers by month of layoff and merges information on all observations in matched employee-employer data for the 3 years before layoff and the 5 years after layoff
  - "Data\_prep\_A3:" prepares sample and variables for post-layoff analysis in relation to formal reemployment outcomes
  - "Data\_prep\_A4:" brings in data from UI registry and prepares variables for post-layoff analysis in relation to UI outcomes
  - "Data\_prep\_A5:" creates last variables useful for all UI-related projects
- B. The code "Data\_prep\_B1" prepares the sample and variables used specifically for the empirical analysis in the paper.

### **3. Stata analysis/Codes for results**

The code generating the graphs in the Stata analysis folder ("Create\_graphs.do") starts by setting the directory and opening a new log file. It then uploads the data and creates in turn:

- The graphs for Figures 2 and 6 (RD graphs);
- The graphs for Figure 4 (histograms of our main outcome on each side of the cutoff);
- The graphs in Figure 5 (full schedule of the statutory UI benefit level).

### **4. R analysis/Codes for results**

The code generating the results in the R analysis folder ("run\_brazil\_qe.R") starts by setting the directory, uploading the data, uploading the R package implementing the methods developed in the paper, and opening a new log file (rdbounds\_r\_output.txt). It then creates in turn:

- The results in Table 2 with the outcome censored at 6 months;
- The results in Table 2 with the outcome censored at 24 months;
- The results in Table 3.

The Stata code ("Create\_figure3.do") takes the results of the "fixed tau" analysis, which is done for the fuzzy case, and create the graphs in Figure 3 using the outcome censored at 24 months.