Replication Package for "Manipulation Robust Regression Discontinuity Designs"

This is a README file for replicating the empirical analyses in the study, titled "Manipulation Robust Regression Discontinuity Designs". The analyses appear in the case study of Section 3.

Overview

Analyses use two datasets from Pop-Eleches and Urquiola (2013a) and Caughey and Sekhon (2011a). Both datasets are down- loaded as their official replication packages. All the analyses are run in a single Rmarkdown file using R. The code produces the numbers in Example 3.1 and two figures in Example 3.2. The code usually runs for less than 3 mins.

Data Availability and Provenance Statements

- ☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

The datasets used in this study are downloaded from openicpsr and Harvard Dataverse as the replication packages for Pop-Eleches and Urquiola (2013a) and Caughey and Sekhon (2011a). Redistribution of both datasets are allowed because the dataset from Pop-Eleches and Urquiola (2013a) is distributed as Creative Commons Attribution 4.0 International Public License; the dataset from Caughey and Sekhon (2011a) is distributed as public domain, CC0 1.0. At this moment, both datasets are available for download, and we include the necessary datasets for our analyses in the replication package.

Statement about Rights

- ☑ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☑ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

License for Data

The dataset from Pop-Eleches and Urquiola (2013a) is licensed under a Creative Commons Attribution 4.0 International Public License. The dataset from Caughey and Sekhon (2011a) is licensed as a public domain, CC0 1.0. See LICENSE.txt for details.

Summary of Availability

- ☑ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Details on each Data Source

Data.Name	Data.Files	Location	Provided	Citation
"Pop-Eleches and Urquiola (2013a)"	data-AER-1.dta	112645-V1/data	TRUE	Pop-Eleches and Urquiola (2013b)
"Caughey and Sekhon (2011a)"	RDReplication.dta	dataverse_files	TRUE	Caughey and Sekhon (2011b)

Data on Pop-Eleches and Urquiola (2013a)

Data on Pop-Eleches and Urquiola (2013a) was downloaded as its replication package from https://doi.org/10.3886/E112645V1. A copy of the data is provided as part of this archive. The data is under a Creative Commons Attribution 4.0 International Public License. The data is attached in the original format without appropriate labeling. We added a codebook `codebook.csv` for the minimal variables used in our analysis. The original data is provided as a proprietary format (dta). We attached its outsheeted format (csv), but the csv file is not used for the analysis; the original proprietary data is readable from a non-proprietary software `R` with the library `haven` as we did in our analysis.

Datafile: `data-AER-1.dta`

Data on Caughey and Sekhon (2011a)

Data on Caughey and Sekhon (2011a) was downloaded as its replication package from https://doi.org/10.7910/DVN/8EYYA2. A copy of the data is provided as part of this archive. The data is in the public domain. The data is attached in the original format without appropriate labeling. We added a codebook `codebook.csv` for the minimal variables used in our analysis. The original data is provided as a proprietary format (dta). We attached its outsheeted format (csv), but the csv file is not used for the analysis; the original proprietary data is readable from a non-proprietary software `R` with the library `haven` as we did in our analysis.

Datafile: `RDReplication.dta`

Code

Code for data cleaning and analysis is provided as part of the replication package. It is also available at https://github.com/SMasa11/replication_mrrd for review. It will be uploaded to the Econometric Society once the paper has been conditionally accepted. The datasets are too large to store directly in GitHub repository and they are stored through `git lfs`. Please contact the maintainer for any questions.

Dataset list

Data file	Source	Notes	Provided
`112645-V1/data/data-AER-1.dta`	openicpsr	Creative Commons 4.0	Yes
`112645-V1/data/data-AER-1.csv`	openicpsr	Creative Commons 4.0	Yes
`dataverse_files/RDReplication.dta`	Harvard Dataverse	Public Domain	Yes

Data file	Source	Notes	Provided
`dataverse_files/RDReplication.csv`	Harvard Dataverse	Public Domain	Yes

Computational requirements

The code, replication_paper.Rmd, runs on macOS 12.6.9 using R 4.3.1 with RStudio. The minimal exterlier packages are loaded at the initial block of the file; nevertheless, the running environment is snapshoted using renv.

Software Requirements

- R 4.3.1
    - `haven` (2.5.3)
    - `rddensity` (2.4)
    - `renv` (0.15.4) is used to generate the lockfile
- RStudio 2023.06.2+561

Controlled Randomness

The procedure does not involve any random number generations.

- ☐ Random seed is set at line _____ of program _____

Memory and Runtime Requirements

For a 2.3GHz Quad-core Intel Core i7 MacBook Pro with 32GB memory, the whole process takes less than three minutes.

Summary

Approximate time needed to reproduce the analyses on a standard (2023) desktop machine:

- ☑ <10 minutes
- ☐ 10-60 minutes
- ☐ 1-2 hours
- ☐ 2-8 hours
- ☐ 8-24 hours
- ☐ 1-3 days
- ☐ 3-14 days
- ☐ > 14 days
- ☐ Not feasible to run on a desktop machine, as described below.

Details

The code was last run on a **4-core Intel-based laptop with MacOS version 12.6.9**.

Description of programs/code

- The program `replication_paper.Rmd` loads two raw datasets from file and compute the density ratio with `rddensity` package.

License for Code

The code is licensed under a GPL-2 license. See LICENSE.txt for details.

Instructions to Replicators

All paths are local. Hence, the desired results in a html file `replication_paper.html` is generated by setting the current directly to the folder where `replication_paper.Rmd` resides and knitting the Rmarkdown file. We recommend replicators to open `RStudio` project `mrrd_replication.Rproj` to conduct the analysis.

List of results and programs

The provided code reproduces:

- ☑ All numbers provided in text in the paper
- ☐ All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below.

Page #	Program	Line Number
Page 24, line 18	replication_paper.Rmd	75
Page 24, line 20	See Pop-Eleches and Urquiola (2013a), Table 5 (2) Panel B	
Page 24, line 24	replication_paper.Rmd	84
Page 26, Figure 3.3 (a)	replication_paper.Rmd	26
Page 26, Figure 3.3 (b)	replication_paper.Rmd	38

References

CAUGHEY, DEVIN AND JASJEET S. SEKHON (2011a): "Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008," Political Analysis, 19 (4), 385–408.

——— (2011b): "Replication data for: Elections and the Regression- Discontinuity Design: Lessons from Close U.S. House Races, 1942-2008," DOI: 10.7910/DVN/8EYYA2.

POP-ELECHES, CRISTIAN AND MIGUEL URQUIOLA (2013a): "Going to a Better School: Effects and Behavioral Responses," American Economic Review, 103 (4), 1289–1324.

——— (2013b): "Replication data for: Going to a Better School: Effects and Behavioral Responses. Nashville, TN: American Economic Association [publisher]," Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-10-11, DOI: 10.3886/E112645V1.