Replication Package for: "Revisiting the Analysis of Matched-Pair and Stratified Experiments in the Presence of Attrition"

Overview

The code in this replication package constructs the analysis of the eight selected papers for empirical application (Groh & McKenzie, 2016; Dhar et al, 2022; Carter et al, 2021; Casaburi & Reed, 2022; Abebe et al, 2021; Hjort et al, 2021; Romero et al, 2020; Annatasio et al, 2020) using R. For each of these papers, we provide `.csv` files (in separate subdirectories) containing the results of our analyses.

Data Availability and Provenance Statements

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.

Details on each Data Source

The paper uses data from 7 RCTs from recent issues of the *American Economic Journal: Applied Economics* and the *American Economic Review*, as well as 1 RCT published in the *Journel of Development Economics*. The availability of all datasets are described below:

- Groh & McKenzie (2016): Data can be downloaded from the published article's main page https://doi.org/10.1016/j.jdeveco.2015.08.003. Copy `MacroinsuranceforMicroentrepreneurs.dta` into the directory `1. Groh & McKenzie (2016)` and run `Data Cleaning.do`.

Datafile: `1. Groh & McKenzie (2016)/MacroinsuranceforMicroentrepreneurs.dta`

- Casaburi & Reed (2022): Data can be downloaded from OPENICPSR https://www.openicpsr.org/openicpsr/project/135021/version/V1/view. Copy the entire `R2` subfolder into the directory `2. Casaburi and Reed (2021)`.

Datafile: `2. Casaburi and Reed (2021)/R2/dta/data_for_analysis/data_bags_cleaned_210630.dta`

- Dhar et al (2022): Data can be downloaded from OPENICPSR https://www.openicpsr.org/openicpsr/project/149882/version/V1/view. Copy the entire `data` subfolder into the directory `3. Dhar, Jain, Jayachandran (2022)`.

Datafile: `3. Dhar, Jain, Jayachandran (2022)/data/bt_analysis_final.dta`

- Carter et al (2021): Data can be downloaded from OPENICPSR https://www.openicpsr.org/openicpsr/project/116761/version/V2/view. Copy the entire `data` subfolder into the directory `4. Carter, Laajaj, Yang (2021)`.

Datafile: `4. Carter, Laajaj, Yang (2021)/data/original/Moz1234panel.dta`

- Abebe et al (2021): Data can be downloaded from OPENICPSR https://www.openicpsr.org/openicpsr/project/127401/version/V1/view. Copy the entire `data` subfolder into the directory `5. Abebe, Caria, Ortiz-Ospina (2021)`.

Datafile: `5. Abebe, Caria, Ortiz-Ospina (2021)/data/generated/MainExperiment_ForAnalysis.dta`

- Hjort et al (2021): Data can be downloaded from OPENICPSR https://www.openicpsr.org/openicpsr/project/122661/version/V1/view. Copy the entire `raw data` subfolder into the directory `5. Abebe, Caria, Ortiz-Ospina (2021)`.

Datafile: `6. Hjort, Moreira, Rao, Santini (2021)/raw data/policy-adoption experiment/policyadoption_raw.csv`

- Romero et al (2020): Data can be downloaded from Harvard Dataverse https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/5OPIYU. Run all necessary Stata code in `PSL Endline Dataverse Files/Analysis/code/StataCode` to obtain the endline dataset, and then save out the endline data as `estimation data.dta` into the directory `7. Romero, Sandefur, Sandholtz (2020)`.

Datafile: `7. Romero, Sandefur, Sandholtz (2020)/estimation data.dta`

- Attanasio et al (2020): Data can be downloaded from the published article's page http://doi.org/10.1257/aer.20150183. Copy the entire `data` subfolder into the directory `8. Attanasio et al (2020)`.

Datafile: `8. Attanasio et al (2020)/data/AER-2015-0183_data_appendix/data/ measures_extended.dta`

Dataset list

Data file	Source	Notes	Provided
`1. Groh & McKenzie (2016)/ MacroinsuranceforMicroentrepreneurs.dta`	JDE Article Main Page	As per terms of use	Yes
`2. Casaburi and Reed (2021)/R2/dta/data_for_analysis/data_bags_cleaned_210630.dta`	OPENICPSR	As per terms of use	Yes
`3. Dhar, Jain, Jayachandran (2022)/data/bt_analysis_final.dta`	OPENICPSR	As per terms of use	Yes
`4. Carter, Laajaj, Yang (2021)/data/original/Moz1234panel.dta`	OPENICPSR	As per terms of use	Yes
`5. Abebe, Caria, Ortiz-Ospina (2021)/data/generated/MainExperiment_ForAnalysis.dta`	OPENICPSR	As per terms of use	Yes
`6. Hjort, Moreira, Rao, Santini (2021)/raw data/policy-adoption experiment/policyadoption_raw.csv`	OPENICPSR	As per terms of use	Yes
`7. Romero, Sandefur, Sandholtz (2020)/estimation data.dta`	Harvard Dataverse	As per terms of use	Yes
`8. Attanasio et al (2020)/data/AER-2015-0183_data_appendix/data/measures_extended.dta`	AER Article Main Page	As per terms of use	Yes

Computational requirements

Software Requirements

The analysis of the results was run using R.

- R 4.2.2
  - tidyverse
  - xtable
  - data.table

Memory and Runtime Requirements

Summary

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

- [ ] < 5 mins

Details

The code was last run on a **8-core Arm-based (Apple M2) laptop with Mac OS**. The entire code took < 5 minutes to run.

Description of programs/code

- Programs in `code/01_simulation` will perform simulation studys on all datasets referenced above and output the results in `data/analysis`. The `Dockerfile` in root directory will run them all.

- `Compiling final results.R` generates Figure 1 in the main text, as well as Table 1 and Table 2. Table 1 and Table 2 results are saved out to the directory `1. Groh & McKenzie (2016)`.

Instructions to Replicators

- Download all replication packages according to instructions above.
- Run `1. Groh & McKenzie (2016)/Data Cleaning.do` to obtain cleaned data.
- From the replication package for `7. Romero, Sandefur, Sandholtz (2020)`, run all necessary Stata code in `PSL Endline Dataverse Files/Analysis/code/StataCode` to obtain the endline dataset, and then save out the endline data as `estimation data.dta` into the directory `7. Romero, Sandefur, Sandholtz (2020)`.
- Open `Compiling final results.R` at the top of the replication package directory. Install the package `pacman` if necessary, then execute the entire code.

List of tables and programs

The provided code reproduces:

- [ ] All tables in the paper

Figure/Table #	Program	Line Number
Table 1	Compiling final results.R	16
Table 2	Compiling final results.R	17
Figure 1	Compiling final results.R	72 - 91

References

Abebe, G., Caria, A. S., and Ortiz-Ospina, E. (2021). The selection of talent: Experimental and structural evidence from ethiopia. American Economic Review, 111(6):1757–1806.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. American Economic Review, 110(1):48–85.

Carter, M., Laajaj, R., and Yang, D. (2021). Subsidies and the african green revolution: direct effects and social network spillovers of randomized input subsidies in mozambique. American Economic Journal: Applied Economics, 13(2):206–229.

Casaburi, L. and Reed, T. (2022). Using individual-level randomized treatment to learn about market structure. American Economic Journal: Applied Economics, 14(4):58–90.

Dhar, D., Jain, T., and Jayachandran, S. (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. American Economic review, 112(3):899–927.

Groh, M. and McKenzie, D. (2016). Macroinsurance for microenterprises: A randomized experiment in post-revolution Egypt. Journal of Development Economics, 118:13–25.

Hjort, J., Moreira, D., Rao, G., and Santini, J. F. (2021). How research affects policy: Experimental evidence from 2,150 brazilian municipalities. American Economic Review, 111(5):1442–1480.

Romero, M., Sandefur, J., and Sandholtz, W. A. (2020). Outsourcing education: Experimental evidence from Liberia. American Economic Review, 110(2):364–400.

Acknowledgements

Some content on this page was copied from Hindawi. Other content was adapted from Fort (2016), Supplementary data, with the author's permission.