README and Guidance

Executive Summary

This file explains how to replicate the results of "Credit, attention, and externalities in the adoption of energy efficient technologies by low-income households" using Stata and R. The two master files run all of the code to generate the data for the figures and tables in the paper. The replicator should expect the code to run for about 120 minutes.

To reproduce all tables and figures, follow the instructions under "Instructions to Replicators".

- The file "List of Tables and Figures.xlsx" provides a comprehensive list of the paper's figures and tables and their sources.
- The file "List of Datasets.xlsx" provides a comprehensive list of data files and their sources.

Data Availability and Provenance Statements

Statement about Rights

We certify that the authors have documented permission to access, use and redistribute/publish the data contained within this replication package.

Summary of availability

Some data cannot be made publicly available.

Details on Data Sources

The data used for this project was collected by the authors in the course of a field experiment, conducted in Nairobi, Kenya. The study is pre-registered on the AEA social science registry (AEARCTR-0002484).

All de-identified data are publicly available. Preserving anonymity requires that we drop from this dataset personal identifiable information including names, date of birth, GPS coordinates, phone numbers, SMSs that contain personal information and payment data. That constraint implies that readers can replicate all our analyses except Figure A2 (for which the readers cannot access data at all, since the figure is directly based on GPS coordinates), and Figures A7, C12 and C13 and Table B3 (which can be only partially replicated, because part of the data cleaning process relies on PII that cannot be made public).

Experiment scripts

SurveyCTO files used to generate the surveys are in folder "Instruments". These work as data dictionaries for the Pilot, Visit 1, Visit 2, Visit 3, and Endline raw data, which can be found in Data/Raw.

Code

Code for data cleaning and analysis is provided as part of the replication package. Code that references PII data has been anonymized, and cannot be executed.

Dataset list

All data used in this paper that does not contain PII is available on the archive in the folder "Data/Raw". There are 6 folders and 2 files inside "Data/Raw".

The file "List of Datasets.xlsx" provides a comprehensive list of data files and their sources.

Folders "Data/Raw/**Pilot**", "Data/Raw/**Visit 1**", "Data/Raw/**Visit 2**", "Data/Raw/**Visit 3**" and "Data/Raw/**Endline**" contain the raw output of the different surveys conducted as part of the experiment. Variables that contain PII were removed from the files in these folders, and are available only in the confidential repository associated to this project. The code to clean these files is provided, but where applicable, lines of code containing PII have been anonymized.

Folder "Data/Raw/**From Confidential**" contains 5 files generated in the confidential repository associated to this project. Some variables that contain PII were removed from these files. In other cases, the variables that contain PII were not deleted but the observations with PII were deleted.

Files "Raw/Data/Visit1_medium_original.dta" and "Raw/Data/Visit1_clean_original.dta" contain data from the Visit 1 survey and the randomized treatments and BDM prices used during Visit 2. These two files were originally created by "MedV1.do" and "CleanV1.do"

- "MedV1.do" requires PII to run: we have included a de-identified version to demonstrate what the code did , but this de-identified version does not run.
- "CleanV1.do" can be run but does not generate the same random assignment as it did when it was originally run. Random assignment was done on a continuous basis as enrolment took place (so that participants could begin receiving SMSes even while others were still being enrolled), and hence the do-file was continuously edited for small errors encountered in the field. This affected the seed draws. When re-run, these files therefore do not generate the random assignments of the files that were eventually used in the project.

All data sources are in one of the following formats: .dta, .csv

Computational requirements

Software Requirements

All code was run using Stata 12 and R 4.1.1.

The exact versions of the Stata packages that are used to generate the results are available in folder "Installers", and are installed by "InstallPackages.do". These packages are:

- ivreg2
- groups
- nmissing
- figout
- renvarsm
- rdrobust
- rdplot

- rdbwselect
- estout
- binscatter
- ranktest
- rscript
- fre
- isvar
- missing
- mdesc
- cfout
- cf3
- irr
- randtreat
- coefplot
- kdens
- moremata
- unique
- findname
- dataout
- whereis
- markstat
- ivqte

The file "Do/setup_R" installs the exact version of all R packages that were used to generate the results. These packages are:

- groundhog
- dplyr
- xtable
- lfe
- ggplot2
- ggthemes
- gridExtra
- corrplot
- tidyr
- plyr
- readxl
- remotes
- extrafont
- car
- fontcm
- Rttf2pt1
- caret

- MASS
- glmnet
- generics

Memory and Runtime Requirements

Approximate time needed to reproduce the analyses on a standard 2021 desktop machine: 2 hours.

Details

When ran using a Macbook Pro 2020 with a 2.3 GHz Quad-Core Intel Core i7 processor and 16GB of RAM, the analysis was completed in two hours.

When ran using a PC with a 1.8 GHz i5-8250U Intel Core i5 processor and 8GB of RAM, the analysis was completed in 2.5 hours.

In both cases, the vast majority of the time was devoted to run the script MLHet.R and the analysis 'ivqte' in script Tables.do.

Description of programs/code

The file "List of Tables and Figures.xlsx" provides a comprehensive list of figures and tables and their sources.

- "1. run_stata.do" sets the root path for the project, creates some empty folders, installs the pre-loaded packages and runs all required Stata scripts sequentially.

- "InstallPackages.do" installs all Stata packages used in the project.

- "0.Master.do" sets the interpretation of all language elements and commands to be the same as it was in Stata 12, and defines macros. This file is called by all the remaining Stata scripts.

- "CleanPilot.do" imports and cleans data from the (2018) pilot. As part of this, the file calls "labels.do", which contains the labels for the different variables.

- "MedV1.do" imports and cleans data from the Visit 1 survey and randomly assigns participants to the different treatments.

- "CleanSMSpreV1.do" creates a list with all the participants that answered SMSs between 2019-04-17 and 2019-06-10, which is used in "CleanV1.do"

- "CleanV1.do" randomly assigns the prices used in Visit 2 during the Becker, DeGroot, Marschak (BDM) mechanism, as well as practice BDM prices and effort tasks (also used in Visit 2).

- "CleanV2.do" imports and cleans Visit 2 data, and creates a file that contains clean data for visits 1 and 2.

- "CleanAllSMS – 1.do" imports all SMSs received during 2019 and removes PII. This an obfuscated version of the script that does this work in the confidential repository. This script does not work in the public repository.

- "CleanAllSMS – 2.do" cleans 2019 SMSs (removes messages without meaningful information and duplicates, converts numbers expressed as words into proper numbers, etc.) and exports them in a format suitable for analysis.

- "CleanV3.do" imports and cleans Visit 3 data, and creates a file that contains clean data for visits 1, 2 and 3.

- "ImportStovesE2020.do" is a SurveyCTO-generated do file that assigns labels to each variable in Endline data.

- "CleanE1.do" imports and cleans data from the 2020 Endline, and creates a file that contains clean data for visits 1, 2 and 3 and the 2020 Endline. This script calls "ImportStovesE2020.do" as part of the process.

- "CleanSMSdataE1 - 1.do" and "CleanSMSdataE1 - 2.do" import and clean raw SMSs sent during 2020.

- "AnalysisData.do" takes as input clean data from visits 1, 2 and 3 and SMS data and prepares it for analysis.

- "AnalysisDataE1.do" merges the output from "AnalysisData.do" with clean data from the the 2020 Endline survey and creates the file that is used to generate most figures, tables and facts.

- "PaymentsData.do" processes the stove repayment data anonymized in the confidential repository. Its output is used in "Figures.do" and "Tables.do" to create Tables X and Y and some in-text numbers.

- "Tables.do" uses data created in previous scripts to generate all the Tables in the paper (see file "List of Tables and Figures" for more detail). This file uses custom functions defined in "Balance_reg.do"

- "Figures.do" uses data create in previous scripts to create all Figures except for A2 (see file "List of Tables and Figures" for more detail).

- "Facts.do" uses output from previous scripts to generate in-text numbers.

- "Figure A12.R" uses data generated in "Tables.do" to create Figure 12 (QTE).

- "PrepForMLTrain.do" prepares the data for use in "MLHet.R"

- "MLHet.R" processes the data generated in "PrepForMLTrain.do" and outputs the data that "FigureA13.R" uses to generate Figure A13 (GATES).

- "Manual_Changes.txt" is a file that explains the rationale for certain data cleaning decisions made during the project. This is not code, it's available only for informative purposes.

.do scripts that contain PII data are commented out in "1. run_stata.do". Those files are public versions of scripts available in the (separate) confidential repository. The versions available in the confidential repository either (i) explicitly contain PII as part of the code, or (ii) make use of variables that contain PII data. The versions available in this public repository (i) do not contain PII in the code (it has been obfuscated), and (ii) can't use PII data, since it has been removed from the raw files. For that reason, the public versions of the scripts do not work. We have shared those public versions so that everyone can see the code used in the private repository.

Instructions to Replicators

1. Edit "1. run_stata.do" at line 6 to adjust the path to that where the project is located in your computer.
2. Run "1. run_stata.do" to run all non-PII Stata code in sequence.
3. Run "2. run_R.R" to run all R code in sequence.

R files use data generated by .do scripts, so "2. run_R.R" must be executed after "1. run_stata.do"

The code reproduces all numbers provided in-text in the paper. All in-text numbers are generated in Facts.do

The code outputs all tables and figures in the paper into the folders /Results/Tables/ and /Results/Figures/ respectively, except for Table 3 and Figure A3.

Table 3 and Figure A3 are available in /Results/Tables/ and /Results/Figures/ respectively, along with additional images and graphics that are not created by code.

Bibliography

Berkouwer, Susanna B., and Joshua T. Dean. 2022. "Data and Code for: Credit, attention, and externalities in the adoption of energy efficient technologies by low-income households." Inter-university Consortium for Political and Social Research [distributor] ID 166661, *American Economic Association [publisher]*. doi:http://doi.org/10.3886/E166661V1.

Berkouwer, Susanna, and Joshua Dean. 2019. "Behavioral barriers to energy efficiency adoption in Kenya: evidence from cookstoves." AEA RCT Registry. July 17. doi:https://doi.org/10.1257/rct.2484.