README File for Dataverse submission of data and do files supporting the paper:

"Financing Municipal Water and Sanitation Services in Nairobi's Informal Settlements"

Version: 5 September, 2023

This replication package generates the main tables and figures used in the paper "**Financing Municipal Water and Sanitation Services in Nairobi's Informal Settlements",** accepted for publication at the Review of Economics and Statistics. The do files include all analysis conducted from original datasets to the creation of the 7 tables in the main paper, together with 19 tables and 3 figures in the online appendix that use data[1]. The datasets provided in this package are the clean, deidentified versions of the study datasets. One master file (*master.do*) runs all the code to generate the data and final tables and figures. The replicator should expect the analysis code to run for about <5 mins, and if including the data preparation and variable construction code for <10 minutes.

A. Instructions

This repository contains all the files necessary to replicate the results shown in the paper. To do so, follow the instructions below.

1. **Download the replication folder** by clicking on the `Access Dataset` button on the main page of the repository.

2. **Use the master do file** to replicate the results in the `do-files' folder. It is only necessary to add your computer's username and path to adjust the default path and set up the working environment to run the replication files. You can select which sections to run by editing the globals at the beginning of the Mater do file. Make sure to run the ***packages*** section to install all necessary packages before running the other sections.

3. **Outputs will be generated when you run the master do file** and saved in the `outputs' folder.

B. Computational requirements

Software requirements

The code was last run with Stata version 14. Required packages include `gweightave', `ietoolkit', `iefieldkit', and `veracrypt'. The ***packages*** section of the master dofile will install all dependencies locally, and should be run once.

Memory and runtime requirements

Approximate time needed to reproduce the analyses on a standard 2018 desktop machine: <10 minutes. The code was last run on a dual-core Intel-based laptop with MacOS version 11.4.

[1] Figure A3 and Figure A4 do not rely on any data. Table A16 (Cost-Benefit Analysis) is produced manually based on inputs from the tables produced in the paper.

C. Description of code

- The code in `dofiles/analysis/final' generates all main tables and figures used in the paper. The code file `master.do' will run them all. Each program called from `master.do' identifies the table or figure it creates (e.g., `table5.do'). Output files are called appropriate names (`table5.tex', `figure1.png') and should be easy to correlate with the manuscript.

- The code in `dofiles/baseline' and `dofiles/endline' and the subfolders `clean' and `construct', cleans the raw encrypted data and constructs the final indicators for each survey round. The do file `master.do' will run them all. Access to encrypted files are required to run cleaning and deidentification code in the `clean' subfolders, and without access the switch for *clean* should be set to zero in `master.do'.

- The code in `dofiles/masterdata' deidentifies master study files. Access to encrypted files are required to run this cleaning and deidentification code, and without access the switch for ***clean*** should be set to zero in `master.do'.

- Ado files have been stored in `dofiles/ado' and the `master.do' files set the ADO directories appropriately.

- Access to the raw utility billing data is restricted. The raw data are cleaned using *billing_bl_clean.do* and *billing_end_clean.do* and the variables are constructed for the analysis using *billing.do*. These .do files create the aggregated and cleaned dataset *billing.dta* which is available in the reproducibility package and can be used to produce all of the tables in the paper. Figures A1 and A5 require the raw disaggregated billing data to be produced. The Master do file allows you to indicate whether you have access to the raw household or billing data and bypasses the code requiring these datasets if you set the respective local "clean" or global "billing_data" to 0. To access the original de-identified/disaggregated billing data please contact the Nairobi City Water and Sewerage Company ICT director Martin Nang'ole to submit a request.

D. List of outputs and code

All analysis code is stored in `dofiles/analysis/main' and `dofiles/analysis/appendix'. All outputs are saved to `outputs'. All the code can be run from the `master.do' script, but the code to recreate each output can also be run independently, as long as the folder globals and custom programs are set using the master script.

Tables (Main Paper)

Table 1: Property owner baseline summary statistics

Table 2: Tenant baseline summary statistics

Table 3: Impacts of engagement on repayment

Table 4: Impacts of targeted enforcement on repayment

Table 5: Spillovers from targeted enforcement

Table 6: Effect of targeted enforcement on owner and tenant outcomes

Table 7: Effect of targeted enforcement on rental market outcomes

Tables (Online Appendix)

Table A1: Correlates with payment behavior

Table A2: Baseline balance

Table A3: Impacts of baseline outreach campaign

Table A4: Study vs. non-study participant sample comparison

Table A5: Differential payment behavior by property owner residence status

Table A6: Differential payment behavior by whether property owner has ever made a payment prior to treatment

Table A7: Differential payment behavior by property-owner-reported financial constraint

Table A8: Differential payment behavior by water access prior to treatment

Table A9: Impact of enforcement on psychological well-being indices

Table A10: Individual variables composing landlord indices

Table A11: Individual variables composing tenant indices

Table A12: Baseline balance (overlap sample)

Table A13: Impacts of engagement on repayment (overlap sample)

Table A14: Impacts of enforcement on repayment (overlap sample)

Table A15: Engagement and enforcement interaction effects (overlap sample)

*Note that Table A16: Cost-Benefit Analysis is produced manually using inputs described in the paper

Table A17: Impacts of contract enforcement on payment (baseline controls)

Table A18: Spillovers of contract enforcement on payment (baseline controls)

Table A19: Impacts of contract enforcement on owner and tenant outcomes (baseline controls)

Table A20: Impacts of contract enforcement on rental market outcomes (baseline controls)

Figures (Online Appendix)

Figure 1: Revenue collection efficiency and number of utility customers

Figure 2: Self-reported reasons for nonpayment

*Note that Figures 3 and 4 are produced manually

Figure 5: Intervention impacts on payment behavior

If you run into any troubles running this code or reproducing results, please contact Aidan Coville at acoville@worldbank.org.