README file for the Replication Package

"Design of Partial Population Experiments with an Application to Spillovers in Tax Compliance"

Data Availability Statement:

For confidentiality reasons, we cannot post any tax microdata used in the analysis.

How to access proprietary data. To access proprietary data, please contact the following employee:

Josefina Currao, *Secretaria de Finanzas y Modernización*

- Municipalidad de Tres de Febrero, Provincia de Buenos Aires, Argentina
- Juan Bautista Alberdi 4840, B1678 Caseros, Provincia de Buenos Aires, Argentina
- Phone: +54 11 4734-2400
- Email: Jcurrao@tresdefebrero.gov.ar or finanzas@tresdefebrero.gov.ar

[A] README file lists all the **replication files** with a brief description of each → we clearly indicate which programs correspond to what results in the paper.

[B] README file also contains a **data dictionary** that defines each variable and gives the data's provenance. "Dictionary.xlsx" defines each variable in the different databases used in the paper.

[C] README file identifies the **version of the software** used (by version number and/or release date) and similarly for the operating system on which the software runs.

- Stata 17 MP Parallel Edition for Windows (64-bit x86-64)
- R version 4.4.2 (2024-10-31 ucrt) -- "Pile of Leaves"
- RStudio 2024.09.1 Build 394 "Cranberry Hibiscus" Release (a1fe401f, 2024-11-03) for windows

Replication files with a brief description of each:

Each line of the do-file is fully commented. We clearly indicate which programs correspond to what results in the paper.

1. **0_master.do** main program that sets the working directory and executes the rest of the Stata codes described below.
   - This do-file sets the Stata command interpreter to **version 16**.
   - The R scripts must be manually run in R.
2. **1_data_baseline.do** Build baseline datasets for Power calculation and Randomization.
   Input file:
   - "Base 1 Cuenta Corriente por Cuenta 2019.dta" dataset with baseline monthly payments in the year 2019.

Output files:

- o "data_illustration.csv" used in R for Table 1. Restricted to street-blocks with more than 7 accounts per street-block.
- o "data_baseline.dta" dataset with baseline payments in the year 2019 restricted to block-streets with 8-50 properties (requested by municipality).

3. **1_assignment.do** randomization code with group-level and individual-level assignment.
   Input files:
   - o "data_baseline.dta" dataset with baseline payments in the year 2019.

   Output file:
   - o "randomization_final.dta" data with the randomized Treatment and Control groups.
   - o **Table A1** with sample sizes for the field experiment. Manually put into "tex" table.
   - o **Table A2** with descriptive statistics in 2019 (baseline year). Manually put into "tex" table.

4. **1_prepare_db.do** prepares the working database to estimate the spillover and total effects.
   Input files:
   - o "Base 1 Cuenta Corriente por Cuenta al 10Mar.dta" is the Main Data (with outcomes and covariates).
   - o "randomization_final.dta" data with the randomized Treatment and Control groups.
   - o "Base total con datos imprenta.dta" data provided by the municipality with additional variables.
   - o "BOLETAS DIGITALES (3).xlsx" flags a few Undelivered Letters.

   Output file:
   - o "working_data.dta" is the clean data we use for the regression analysis.

5. **1_balance_checks.do** balance check regression analysis comparing treatment and control.
   Input file:
   - o "working_data.dta"

   Output file:
   - o "saturated_regressions_balance.tex" (**Table A3** in the appendix).

6. **1_Figure_1.do** Creates **Figure 1** of the paper: Distribution of cluster sizes in six partial population experiments.

7. **1_Table_A6.R** (note: script must be run in R) Creates **Table A6** of the paper. Standard errors and MDEs using the cluster size distribution of four studies. The code considers three values for the intraclass correlation (rho_00). It must be manually changed to 0.1, 0.5, or 0.8.

8. **1_Figure_2.R** (note: script must be run in R) Creates **Figure 2** of the paper: Power functions - numerical illustration.

9. **1_Table_1.R** (note: script must be run in R) Produces **Table 1** of the paper: Constrained and optimal cluster assignment probabilities and MDEs for our binary outcome of interest.
   Input file:
   - o "data_illustration.csv"

   Output:
   - o Matrix "Results_R": MDEs and restricted cluster probabilities (**Table 1**).
   - o Matrix "Results": MDEs and optimal cluster probabilities (**Table 1**).

10. **2_analysis_cuota10_rawgraphs.do** Distribution of payment date by T vs C; Cumulative payment rate by calendar date.
    Input file:

- o "working_data.dta"

Output files:
- o **Figure A5:** "payments-density1.pdf" and "payments-density2.pdf" Distribution of payment date for treated, untreated, and pure control.
- o **Figure B6**. Payment rates: Treated groups vs Pure control blocks. Breaks treated units into subgroups.
  - Top panel "payment-rate-Oct2020-levels-treated-subgroups.pdf"
  - Bottom panel "payment-rate-Oct2020-TE-subgroups.pdf"
- o **Figure B7**. Payment rates: Untreated groups vs Pure control blocks. Breaks untreated units into subgroups; pools treated units:
  - Top panel "payment-rate-Oct2020-levels-untreated123.pdf"
  - Bottom panel "payment-rate-Oct2020-TE.pdf"

11. **2_analysis_cuota10.do** estimates total and spillover effects for the intervention bill (October 2020) and a pre-intervention bill (September 2020).

Input file:
- o "working_data.dta"

Output files:
- o **Figures B.15, B.16 and B.17** "share_Nbills_indiv_2019.pdf" "share_compliance_2019.pdf" "share_compliance_2018_2019.pdf" "share_compliance_2018_2020.pdf" and "share_compliance_transition.pdf"
- o **Table 2** "saturated_regressions_REStat.tex"
- o **Table A4** "saturated_regressions_REStat_placebo.tex" Direct and spillover effects for the pre-intervention September 2020 bill. Pooling observations and splitting below and above median compliance in 2019.
- o **Figure 3**
  - Top panel: "payment-rate-diff-treated3-notitle.pdf" and "payment-rate-diff-untreated3-notitle.pdf"
  - Middle panel: "payment-rate-diff-treated3-abovep50-notitle.pdf" and "payment-rate-diff-untreated3-abovep50-notitle.pdf"
  - Bottom panel: "payment-rate-diff-treated3-belowp50-notitle.pdf" and "payment-rate-diff-untreated3-belowp50-notitle.pdf"
- o **Figure B8**. Six figures in top, bottom, and middle panels: payment-rate-diff-treated1.pdf payment-rate-diff-treated2.pdf payment-rate-diff-treated3.pdf payment-rate-diff-untreated1.pdf payment-rate-diff-untreated2.pdf payment-rate-diff-untreated3.pdf

12. **2_analysis_cuota9.do** Produces the figures with the point estimates of total and spillover effects for the pre-intervention bill (September 2020). For Table A.4 see **2_analysis_cuota10.do.**

Input file:
- o "working_data.dta"

Output files:
- o **Figure B.9** payment-rate-cuota9-diff-treated1.pdf payment-rate-cuota9-diff-treated2.pdf payment-rate-cuota9-diff-treated3.pdf payment-rate-cuota9-diff-untreated1.pdf payment-rate-cuota9-diff-untreated2.pdf payment-rate-cuota9-diff-untreated3.pdf

13. **3_analysis_paperless.do** Results for Section B.2.1 on subscription to electronic billing.

Input files:

- o "suscripciones_20210308.xlsx" Subscriptions to paperless electronic billing
- o "randomization_final.dta" data with the randomized Treatment and Control groups.
- o "Base 1 Cuenta Corriente por Cuenta al 10Mar.dta" Merge with main data to get block-id for standard errors

Output files:
- o **Figure B.10** subscription-rate-DiD-treated1.pdf subscription-rate-DiD-treated2.pdf subscription-rate-DiD-treated3.pdf subscription-rate-DiD-untreated1.pdf subscription-rate-DiD-untreated2.pdf subscription-rate-DiD-untreated3.pdf
- o **Table A.5** "saturated_regressions_ebill_2.tex"

14. **3_analysis_bills.do** Results for Section B.2.2 on backward and forward payments (cancel past-due debt and improve payment likelihood of subsequent bills).

Input file:
- o "working_data.dta"

Output files:
- o **Figure B.11** Level-probpay.pdf Level-probpay-timely.pdf DinD-probpay.pdf DinD-probpay-timely.pdf

15. **4_analysis_cuota10_buffer.do** Robustness check to analyze whether untreated blocks are affected by the intervention.

Input files:
- o "working_data.dta"
- o "cuadras_buffers.dta" street-block-level database containing treatment intensity for buffers of 100, 200, 300 meters around the centroid of each street-block. Built by staff from the *Tres de Febrero* municipality using GIS tools.

Output files:
- o **Figure B.12.**
  - Top panel: is an illustration.
  - Middle panel: Exposure distribution (share of treated units around pure control blocks). buffer_share_T100.pdf buffer_share_T200.pdf buffer_share_T300.pdf
  - Bottom panel: payment rates for pure controls as a function of exposure. buffer_T0_100.pdf buffer_T0_200.pdf buffer_T0_300.pdf
- o **Figure B.13.** Payment rates and exposure of untreated blocks above and below median 2019 compliance, 100 meters buffer. buffer_T0_100_pool.pdf buffer_T0_100_abovep50.pdf buffer_T0_100_belowp50.pdf
- o **Figure B.14.** Payment rates and exposure of untreated blocks and blocks with 80% treated units, 100 meters buffer. buffer_T0T3_100.pdf buffer_T0T3_100_aug.pdf buffer_T0T3_100_jul.pdf buffer_T0T3_100_jul.pdf

Figure A.2. (map) constructed by municipality employees using GIS tools and the data with randomized assignment "randomization_final.dta".