Exploring the Gender Divide in Real Estate Ownership and Property Tax Compliance

Code Repository Documentation

Tatiana Flores*
Guillermo Cruces†
Jose Carlo Bermúdez ‡
Thiago Scot§
Juan Luis Schiavoni¶
Dario Tortarolo‖

*DIME, World Bank. Email: rfloresibarra@worldbank.org
†Nottingham & CEDLAS-UNLP. Email:guillermo.cruces@nottingham.ac.uk
‡PUC-Chile. Email: jcbrmudez@uc.cl
§DIME, World Bank. Email: tscot@worldbank.org
¶DECRG, World Bank. Email: jschiavoni@worldbank.org
‖Corresponding author. DECRG, World Bank. Email: dtortarolo@worldbank.org

Contents

1 Overview

The codes in this repository replicate the tables and figures from "Exploring the Gender Divide in Real Estate Ownership and Property Tax Compliance", by Flores, Cruces, Bermúdez, Scot, Schiavoni, and Tortarolo. The replication folder contains the codes to go from the raw administrative data to the results in the paper.

This documentation is structured as follows. Section 2 describes the data sources and their availability. Section 3 describes the datasets used in the analysis. Section 4 details the computational requirements. Section 5 provides instructions to replicators. Section 6 provides a mapping between the codes and the tables and figures of the paper. Finally, section 7 describes the codes, including data creation codes (section 7.1), analysis codes (section 7.2), and programs (section 7.3).

2 Data Availability and Provenance Statements

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

2.1 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.

2.2 Summary of Availability

☐ All data **are** publicly available.

☐ Some data **cannot be made** publicly available.

☑ **No data can be made** publicly available.

The administrative data for this project is owned by the local revenue agency in *Tres de Febrero*. The data were made available to us exclusively for this research project through collaboration agreements between the corresponding author and the government agency.

While at the time of the writing, we cannot share the data for other projects, we can provide access for replication purposes of this study (conditional on the signing of a confidentiality agreement and a security agreement). Individuals interested in accessing the data for replication purposes can contact Dario Tortarolo (dtortarolo@worldbank.org). We will provide reasonable assistance to requests for clarification and replication. Researchers interested in obtaining the proprietary data for their own analyses can contact the agency directly at finanzas@tresdefebrero.gov.ar.

NOTE. In cases when de-identified datasets are provided by the authors for replication purposes, given their sensibility and according to The World Bank guidelines regarding access to information and personal data security, all files containing de-identified records must be immediately deleted after replication exercises.

2.3 Details on each Data Source

The analysis is based on de-identified administrative raw data provided by the local revenue agency in *Tres de Febrero*. Raw data is available mainly from 2018 to 2024 and includes:

- **Payments data.** This data includes all the registers of payments made by taxpayers (property owners) in *Tres de Febrero*. The data contain the following billing details: account number (unique property ID), address, name of locality (neighborhood), year and month of the bill (12 bills per year), monthly fee (in pesos), paid fee (amount in pesos), due date, date of payment, days overdue, means of payment (cash or electronic), type of account (residential, retail store, manufacturer), and linear front meters of the lot/property.
- **Property ownership, assessed value, and gender.** Tres de Febrero receives access to administrative data from the Province of Buenos Aires containing the assessed values of local properties and individual tax identifiers of up to two owners.
- **RCT data.** We have access to treatment assignments from a large-scale randomized communication campaign conducted by Cruces, Tortarolo, and Vazquez-Bare (2023)[1], who estimate direct and spillover effects on property tax compliance.

1. *Design of Partial Population Experiments with an Application to Spillovers in Tax Compliance.*

3 Datasets

This section describes every dataset used to replicate the results. Table 1 displays the set of raw datasets (found in `Raw` data folder of the replication package) needed for full replication from scratch. Table 2 enlists the set of processed datasets (generated by code and stored in the `Processed` folder of the replication package). Those non-available files for public disclosure are flagged with a "No" in the fourth column, while publicly available datasets provided in the replication repository are flagged with a "Yes". The associated data category is displayed in the fifth column, while the last column describes the approximate date the raw data was accessed for the first time.

Table 1: Raw datasets

Dataset	**Description**	**Notes**	**Provided**	**Category**	**1st Time Accessed**
3df.dta	Working shapefile at the cadastral level	Confidential	No	Property ownership, assessed value, and gender.	August, 2024
3df_manzana.dta	Working shapefile including property identifier and coordinates at the block level.	Confidential	No	Property ownership, assessed value, and gender.	August, 2024
3df_manzana_shp.dta	Working shapefile at the manzana level.	Confidential	No	Property ownership, assessed value, and gender.	August, 2024
3df_shp.dta	Shapefile of Tres de Febrero	Confidential	No	Property ownership, assessed value, and gender.	August, 2024
cadastro.dta	Dataset at the property level including the account number and address of the property.	Confidential	No	Property ownership, assessed value, and gender.	January, 2024
db_gender.dta	Working dataset identifying the gender of each property owner at the account level.	Confidential	No	Property ownership, assessed value, and gender.	January, 2024
db2_170224f.dta	Dataset at the monthly level between 2018-2024 including payments.	Confidential	No	Payments data.	February, 2025
edad_cuit.dta	Working dataset including code for birth year of every owner.	Confidential	No	Property ownership, assessed value, and gender.	January, 2024
working_data_3dF_all.dta	Dataset at the monthly level between 2018-2021, including payments, cadaster assessed property values, and dwelling characteristics.	Confidential	No	Property ownership, assessed value, and gender.	January, 2024
working_data_RCT_Oct20.dta	Working dataset including results of the RCT of 2020.	Confidential	No	RCT data.	January, 2025

Table 2: Processed datasets

Dataset	**Description**	**Notes**	**Provided**
3df_manzana_new.dta	Cleaned version of 3df_manzana.dta	Confidential	No
3df_new.dta	Cleaned version of 3df.dta	Confidential	No
cadaster_seccion_level_data_map.dta	Cleaned version of 3df.dta including compliance outcomes at the cadastral level	Confidential	No
generic	Dataset containing generic numbers (número de cuenta) for observations with missing account numbers.	Confidential	No
invariant_property_traits.dta	Time invariant traits at the property level.	Confidential	No
manzana_level_data_map.dta	Working shapefile combining information at the block level and valuation of properties.	Confidential	No
owed_2019.dta	Data collapsing the taxes owed in 2019 by type of owner (male, female, co-owned).	Confidential	No
p99.dta	Database includes the assessed value percentile at the year-account number level.	Confidential	No
p99_9.dta	Database including the G99 percentiles (splitting the top 1% into 10 bins) of assessed value at the year-account number level.	Confidential	No
str.dta	Dataset including Statutory tax rates for the period 2018-2020 as in *Ordenanza Impositiva*.	Confidential	No
working_data_3dF_yearly.dta	Working dataset at the year level including most of the outcomes as in working_data_3dF_all.dta, but also includes relevant outcomes such as effective tax rates, and deciles/percentiles for the distribution of property assessed values.	Confidential	No

4 Computational Requirements

4.1 Software requirements

Required software is `Stata`. The code was run in version 18, but it perfectly works in any previous version of `Stata`. The main difference between `Stata 18` and older versions is the appearance of graphs. Our master do file includes a global setting that harmonizes the figure's appearances whether the code is run in version 18 or any other previous one.

4.2 Memory and runtime requirements

The code was run for the last time in a Thinkpad/Lenovo E16 laptop with Windows 11 pro version 23H2, 32 GB RAM, and an AMD Ryzen 5 7535HS processor with Radeon Graphics 3.3 GHz. A replicator could expect the whole code to run in ≈23 minutes with those settings.

5 Instructions to replicators

5.1 Folder structure

For full replication of the project, we suggest the following, complete folder structure. Replications of the analysis only starting with the cleaned project datasets should mirror `2_codes` and `3_analysis`.

- `1_Data`
  - `1_Raw`
  - `2_Processed`
- `2_Codes`
- `3_Analysis`
  - `1_Figures`
  - `2_Tables`

We provide the do-files to create the datasets used in the project and the subsequent analysis. In the master do-file, we set the `paths` for every so the above folder tree can run easily. The codes described in section 7 rely on this folder structure.

6 List of tables, and figures

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.

6.1 Mapping of tables and analysis code

Table 3 provides a mapping between all tables of the paper (including the online appendix) and the codes producing these results. The codes generate .tex files containing the tables.

Table 3: Mapping of tables and analysis code

Exhibit name	**Output filename**	**Code**
Table 1	`statutory_tax_rates.tex`	`2a_analysis_tables.do`
Table 2	`descriptive_property_year.tex`	`2a_analysis_tables.do`
Table 3	`reg_payments.tex`	`2a_analysis_tables.do`
Table 4	`regressions_timely_payments.tex`	`2c_analysis_RCT.do`
Table A1	`_`	`Own elaboration`
Table A2	`regressions_balance.tex`	`2c_balance_checks.do`

6.2 Mapping of figures and analysis code

Table 4 provides a mapping between all figures of the paper (including the online appendix) and the codes producing these results. The codes generate files in .pdf format containing the graphs.

Table 4: Mapping of figures and analysis code

Exhibit Name	Panel	Output filename	Code
Figure 1		map_seccion_catastral_properties_compliance.png	3b_maps_seccion_cadastro.do
Figure 2		ownership_scatter_gender.pdf	2b_analysis_figures.do
Figure 3		cdf_assessed_value_gender.pdf	2b_analysis_figures.do
Figure 4		bar_taxes_gender.pdf	2b_analysis_figures.do
Figure 5		prob_pay.pdf	2b_analysis_figures.do
Figure 6		effective_tax_rate.pdf	2b_analysis_figures.do
Figure 7		etr_ownership.pdf	2b_analysis_figures.do
Figure 8	a,b	Level_probpay_timely_control.pdf Level_probpay_control.pdf DinD_probpay_timely_gender.pdf DinD_probpay_gender.pdf	2d_analysis_RCT.do
Figure 9	a,b	payments_density_Oct20.pdf payment_rate_diff_Oct20_gender.pdf	2e_analysis_cuota10.do
Figure 10	a-b	Level_probpay_timely_control_quintiles.pdf DinD_probpay_timely_gender_quintiles.pdf	2d_analysis_RCT.do.do
Figure A1		map_mzn_res_property_number_arbavaluation_2021.png	3a_maps.do
Figure A2		hist_assessed_value_2018.pdf	2b_analysis_figures.do
Figure A3		residential_type_percentile.pdf	2b_analysis_figures.do
Figure C4		charges_deciles_2019.pdf.pdf	2b_analysis_figures.do
Figure C5		time_series_payment_owner.pdf	2b_analysis_figures.do
Figure C6		prob_annual_payment_gender.pdf	2b_analysis_figures.do
Figure C7		Carta3dF_ej.pdf	RCT documentation
Figure C8	a,b	payments_density_Jul20.pdf payment_rate_diff_Jul20_gender.pdf	2e_analysis_cuota10.do
Figure C9	a,b	payments_density_Jul20.pdf payment_rate_diff_Jul20_gender.pdf	2c_analysis_RCT.do

7 Description of programs and code

The project is organized as follows. First, we use the raw data files to create the datasets used in the analysis. These codes are described in section 7.1. Second, we describe the data analysis code in section 7.2. Some of the code relies on programs downloaded from the Statistical Software Components (SSC) archive. These programs are described in section 7.3.

Global paths to the different folders of the project (input files, output tables, figures, etc.) are established in the master do-file. The replicator can centrally adjust the links in these files without adjusting the (relative) paths in the specific cleaning or analysis parts. Moreover, the master do-file installs all packages needed in the project that are downloaded from SSC.

7.1 Creation of data sets for analysis

- 1b_prepare_variables.do
  Harmonize the working dataset, creates additional variables to get inequality measures, and aggregate at the yearly level.

  Inputs :
  $raw_data/working_data_3dF_all.dta
  $raw_data/edad_cuit.dta
  $raw_data/db_gender.dta

  Output:
  $processed_data/invariant_property_traits.dta
  $processed_data/generic.dta
  $processed_data/p99.dta
  $processed_data/p99_9.dta
  $processed_data/working_data_3dF_yearly.dta

7.2 Analysis

- 2a_analysis_tables.do
  Creates tables with summary statistics on property and gender.

  Input:
  $processed_data/working_data_3dF_yearly.dta
  $processed_data/invariant_property_traits.dta
  $raw_data/db_gender.dta

  Output:
  $processed_data/str.dta
  $Tables/statutory_tax_rates.tex
  $Tables/descriptive_property_year.tex
  $Tables/reg_payments.tex

- 2b_analysis_figures.do
  Creates figures describing data on property taxes and gender.

  Input:
  $processed_data/working_data_3dF_yearly.dta
  $processed_data/invariant_property_traits.dta
  $raw_data/db_gender.dta
  $raw_data/db2_170224f.dta

  Output:
  $processed_data/owed_2019.dta
  $Figures/time_series_payment_owner.pdf
  $Figures/bar_taxes_gender.pdf
  $Figures/charges_deciles_2019.pdf
  $Figures/hist_assessed_value_2018.pdf
  $Figures/cdf_assessed_value_gender.pdf
  $Figures/effective_tax_rate.pdf
  $Figures/etr_ownership.pdf
  $Figures/prob_pay.pdf
  $Figures/prob_annual_payment_gender.pdf
  $Figures/residential_type_percentile.pdf
  $Figures/ownership_scatter_gender.pdf

- 2c_balance_checks.do
  Creates tables with balance checks between treatment and control groups.

  Input:
  $raw_data/working_data_RCT_Oct20.dta
  $raw_data/db_gender.dta

  Output:
  $Tables/regressions_balance.tex

- 2d_analysis_RCT.do
  Main analysis from RCT.

  Input:
  $raw_data/working_data_RCT_Oct20.dta
  $raw_data/db_gender.dta

  Output:
  $Figures/Level-probpay-timely-control.pdf
  $Figures/Level-probpay-control.pdf
  $Figures/DinD-probpay-gender.pdf
  $Figures/DinD-probpay-timely-gender.pdf
  $Figures/DinD-probpay-timely-gender-quintiles.pdf
  $Figures/Level-probpay-timely-control-quintiles.pdf
  $Figures/DinD-probpay-timely-gender-quintiles-placebo.pdf
  $Figures/Level-probpay-timely-control-quintiles-placebo.pdf
  $Tables/regressions_timely_payments.tex

- 2e_analysis_cuota10.do
  Analysis of bill for October 2020

  Input:
  $raw_data/working_data_RCT_Oct20.dta
  $raw_data/db_gender.dta

  Output:
  $Figures/payments-density-Oct20.pdf
  $Figures/payment-rate-diff-Oct20-gender.pdf
  $Figures/payments-density-Jul20.pdf
  $Figures/payment-rate-diff-Jul20-gender.pdf

- 3a_maps.do
  Maps at the block level.

  Input:
  $raw_data/3df_manzana.dta
  $raw_data/cadastro.dta
  $raw_data/3df_manzana_shp

  Output:
  $processed_data/3df_manzana_new.dta
  $processed_data/manzana_level_data_map.dta
  $Figures/map_mzn_res_property_number_arbavaluation-2021.png

---

- 3b_maps_seccion_cadastro.do
  Plot maps for the properties number and average values at the seccion cadastral and manzana level

  Input:
  $raw_data/3df.dta
  $processed_data/working_data_3dF_yearly.dta
  $raw_data/cadastro.dta
  $raw_data/3df_shp

  Output:
  $processed_data/cadaster_seccion_level_data_map.dta
  $processed_data/3df_new.dta
  $Figures/map_seccion_catastral_properties_compliance.png

7.3 Programs

Our code utilizes the following commands installed via SSC manually through the master do-file:

- `binscatter`
  Michael Stepner, 2013. "BINSCATTER: Stata module to generate binned scatterplots," Statistical Software Components S457709, Boston College Department of Economics, revised 24 Nov 2013.

- `eststo`
  Ben Jann, Institute of Sociology, University of Bern, jann@soz.unibe.ch

- `estout`
  Ben Jann, Institute of Sociology, University of Bern, jann@soz.unibe.ch

- `freqindex`
  Julio Raffo, 2015. "FREQINDEX: Stata module to generate an index of terms from string variable", Statistical Software Components S457993, Boston College Department of Economics, revised 13 Apr 2019.

- `geoplot, geoframe`
  Jann, B. (2023). geoplot: Stata module to draw maps. Available from https://ideas.repec.org/c/boc/bocode/s459211.html.

- `matchit`
  Julio Raffo, 2015. "MATCHIT: Stata module to match two datasets based on similar text patterns", Statistical Software Components S457992, Boston College Department of Economics, revised 20 May 2020.

- `moremata`
  Ben Jann, 2005. "MOREMATA: Stata module (Mata) to provide various functions", Statistical Software Components S455001, Boston College Department of Economics, revised 24 Nov 2022.