# README: Electronic Payment Technology and Tax Compliance: Evidence from Uruguay's Financial Inclusion Reform

This file explains how to replicate the analysis for the paper "Electronic Payment Technology and Tax Compliance: Evidence from Uruguay's Financial Inclusion Reform" (American Economic Journal: Economic Policy, by Anne Brockmeyer and Magaly Saenz Somarriba), using STATA.

## Data Availability Statement

### Administrative data

Access to the confidential tax data for Uruguay (DGI Uruguay, 2009-2016)

Most of the analysis in this paper uses firm-level VAT declarations and firm characteristics from the taxpayer registry held by the Uruguayan tax authority. The authors accessed this data under the requirement that no micro data could be shared with third parties. The data that the research team obtained is stored on a secure World Bank server which only authorized World Bank staff and consultants can access. The Uruguayan Tax Administration does not currently have a formal policy for granting access to these data. Any other person wishing to obtain access to the data would therefore have to approach the director of the Tax Administration and the head of the economic studies department and make a business case to them for access to the data. The case may also be reviewed by the Ministry of Finance. Since March 2020, the authors have not been able to obtain access to additional data for this project (beyond the data they had already accessed previously). The Tax Administrations' policies for granting data access to researchers may change in the future, e.g. with a change in government.

The authors are happy to facilitate introductions for researchers proposing projects for which they could make a business case. The authors are also committed to run any sensitivity analyses and share the results with anyone wishing to run such analyses.

A replication analysis via the World Bank DIME replication service is being carried out.

If the AEA data editor seeks to access the data, or seeks to nominate a third-party researcher to conduct an audit for the replication files, the authors may be able to provide them access to the data through the World Bank secure data server, subject to some conditions. In particular, the individual seeking to access the data would either already be a World Bank staff or consultant, or needs to be provided a non-fee consulting contract with the World Bank, which is subject to World Bank management approval. The individual seeking to access the data will also need to sign and abide by the World Bank's data confidentiality agreement.

Access to the confidential tax data for Costa Rica (DGT Costa Rica, 2013)

Appendix Figure D.4 (share of card sales in total reported sales in Costa Rica) uses micro administrative data from the Costa Rican tax administration as in the paper "Taxation, Information and Withholding: Evidence from Costa Rica" by Anne Brockmeyer and Marco Hernandez: https://www.annebrockmeyer.com/uploads/1/2/1/4/121485108/costa_rica_paper_2022 -11-14.pdf. Specifically, the data uses the D104 sales tax declaration and the D153 credit and debit card withholding form for the year 2013. From the D104 data, we use the tax return boxes 20, 21, 22 and 23 (variables: ventas por exportaciones, ventas exentas y authorizadas sin impuesto, ventas gravadas, otros rubros a incluir en la base imponible) and sum them to construct total sales. From the D153 declaration, we use the variable "monto total de transacciones", which captures the total amount of card transactions. The access to this data is very similar to the access for the Uruguay data described above. The data is owned by the Costa Rican Tax Administration. The authors accessed this data under the requirement that no micro data could be shared with third parties. The data that the research team obtained is stored on a secure World Bank server which only authorized World Bank staff and consultants can access. The Costa Rican Tax Administration does not currently have a formal policy for granting access to these data. Any other person wishing to obtain access to the data would therefore have to approach the director of the Tax Administration and make a business case to them for access to the data. The case may also be reviewed by the Ministry of Finance.

The authors are happy to facilitate introductions for researchers proposing projects for which they could make a business case. The authors are also committed to run any sensitivity analyses and share the results with anyone wishing to run such analyses. A replication analysis via the World Bank DIME replication service is being carried out. If the AEA data editor seeks to access the data, or seeks to nominate a third-party researcher to conduct an audit for the replication files, the authors may be able to provide them access to the data through the World Bank secure data server, subject to the above-listed conditions.

*Publicly available data*

The following datasets need to be downloaded and saved in the folder _DATA_CODES. World Bank global findex data (Figure A1, panels A.I and A.II, and Figure A.9) (Demirguc-Kunt et al., 2011 and 2017). The Findex data can be downloaded here: https://www.worldbank.org/en/publication/globalfindex/Data (go to country-level data and download the stata version). The codes merge this data with GDP per capita data from the World Development Indicators (World Bank, 2019) (also used in some of the other figures mentioned below). To download the GDP per capita data, go to https://databank.worldbank.org/source/world-development-indicators, and select: Countries: all, time period: 2011 and 2017, series: GDP per capita, constant 2015 US$ Unselect everything else. Then download as a CSV file. Unzip it, and discard the smaller of the two resulting CSV files (which is a meta data file). Save the other file under the name WDI_gdp_pc.csv in the data folder.

Informality rates (Figure A1, panel B) (Medina and Schneider, 2018)

The data is from Table A.1. in the IMF working paper by Schneider and Medina (2018): Shadow economies around the world: What did we learn over the last 20 years?

https://www.imf.org/en/Publications/WP/Issues/2018/01/25/Shadow-Economies-Around-the-World-What-Did-We-Learn-Over-the-Last-20-Years-45583 We extracted the data from the pdf into a dta file (data_informality.dta) which is part of the replication package (in the folder PUBLIC_DATA_CODES – to be transferred into the data folder).

Uruguay households finance survey (Figure D.5, panels A and C) (Universidad de la Republica Uruguay, 2017)

These data, used to construct figure D5, panels A and C, can be downloaded here: https://docs.google.com/forms/d/e/1FAIpQLSdJqJG-3gH2ykI_woUf8Xx4hl47bI_HJTDFM3QZSeNkTM9-Pw/viewform?embedded=true If this does not work, go the his link: https://cienciassociales.edu.uy/departamento-de-economia/investigacion/encuesta-financiera-de-los-hogares-uruguayos/ Go to EFHU 2, and click "solicitor accesso a microdatos". A form will open that you should fill in and submit. The data will then be sent to the email address you indicate when submitting the form.

ENIGH household survey data from the Uruguay National Institute of Statistics (INE) (Figure D.5, panel B) (National Institute of Statistics, 2005-2006)

To obtain these data, go to the following link: https://www4.ine.gub.uy/Anda5/index.php/catalog/42, click on the tab "obtener microdatos", register as a user and then enter your account details to enter the protected part of the website. Then download "Conjunto de datos de los hogares", "Características demográficas de las personas del hogar" and "Consumos". Unzip the files. You should obtain the files personas.DBF, hogar.DBF and consumos.DBF. Add these to the data folder. Alternatively (or if the above website link is not working), you can send an email to difusion@ine.gub.uy to request the micro data. In the document "Data request Spanish for INE.docx" in this replication folder, you will see the project description we provided to INE. Note that we used the 2005-2006 data, as it was the latest available data when we conducted the analysis. The 2016 data has recently been released.

Argentina Central Bank Means of Payment Statistics (Figure B.8) (Argentina Central Bank, 2022)

The data needs to be downloaded from: https://www.bcra.gob.ar/PublicacionesEstadisticas/Cuadros_estandarizados_series_estadisticas.asp selecting the following options: –> Volumen de transacciones, instrumentos de pago, infraestructura del sistema financiero e informaciones por división política.–> Medios de pago –> Volumen de transacciones e instrumentos de pago, por tipo de depósito y canal. This should download an excel file titled "debser". Add this file to the data folder.

Data References

- Argentina Central Bank, "Means of Payment Statistics," 2022. Accessed: 2022-12-04.
- Demirguc-Kunt, Asli, Leora Klapper, Dorothe Singer, Saniya Ansar, and Jake Hess, "Global Findex Database," 2011 and 2017. Accessed: 2023-11-01.
- DGI Uruguay, "Administrative Tax Records," 2009-2016. Accessed: 2015-2017

- DGT Costa Rica, "Administrative Sales Tax Records," 2013. Accessed: 2015
- Medina, Leandro and Friedrich Schneider, "Shadow economies around the world: what did we learn over the last 20 years? Manual data extraction from table A.1," IMF Working Papers, 2018.
- National Institute of Statistics, "Household Income and Expenditure Survey," 2005-2006. Accessed: 2022-12-04.
- Universidad de la Republica Uruguay, "National Household Finance Survey," 2017. Accessed: 2022-11-13.
- World Bank, "World Development Indicators," 2019.

## Computational Requirements

TECHNOLOGY ENVIRONMENT

The public part of the code was reproduced in a computer with the following specifications: - OS: Windows 11 Enterprise - Processor: Intel(R) Core(TM) i5-1145G7 CPU @ 2.60GHz - Memory available: 15.7 GB - Software version: Stata version 17

The private part of the code was reproduced in a private server with the following specifications: - OS: Windows 10 Enterprise - Processor: Intel(R) Xeon (R) CPU @ 2.80GHz - Memory available: 71.1 GB - Software version: Stata version 17

## Instructions for Data Preparation and Analysis

This replication package contains two sets of codes. The first set of codes, in the folder ADMIN_DATA_CODES, draws on the confidential tax data. For users replicating the analysis on the World Bank secure tax data server, this folder should be transferred into the server (into the folder "Replication") via OneDrive or with the help of World Bank IT (Brandon Mills). The first few lines in the Master do-file tell you which file paths you need to update and which folders you need to create so that the code executes correctly. Then please execute the master file, and this should run all other codes.

Some of the do files execute more than one exhibit, for this we created globals to select which specific to execute. To replicate everything make sure all options of globals in lines 80, 91, 108, 191-192, and 214 are all un-commented out.

At the end, one manual step is required: For the figures did_2lines_nliab_ab_10_11_log_modif_99_r7_as1_esp10_wholesale.gph and did_2lines_nliab_ab_10_11_log_modif_99_r99_as1_esp10_wholesale.gph [figure 3, panel C.I, and figure C.2, panel C.I], open the gph files, and in the note below the graph, add the point estimate (PE) and standard error (SE) from the synthetic difference-in-difference (SDID) estimation. For figure did_2lines_nliab_ab_10_11_log_modif_99_r7_as1_esp10_wholesale.gph [figure 3, panel C.I0], the PE and SE are found in 1 did_robust_table_logmodif_mixed.tex, column 9 [table 1]. For Figure did_2lines_nliab_ab_10_11_log_modif_99_r99_as1_esp10_wholesale.gph [figure C.2, panel C.I], the PE and SE are found in table did_robust_table_logmodif_mixed_r99.tex, column 9 [table C.7]. After changing the PE and SE manually in these figures, export the figure as PDF, adding the postscript _edit to the name. Note that we opted for having this

manual step, as it allows us to otherwise run the SDID separately from the main DID analysis, which facilitates trouble-shooting the code, as the SDID takes a long time to run.

The second set of codes, in the folder PUBLIC_DATA_CODES, runs on publicly available data described above, and should not be transferred into the secure data server. These codes produce only appendix exhibits. The first few lines in Master_public_data.do tell you which file paths you need to update so that the code executes correctly. If any of the raw datasets are compressed, please unzip them.

### Execution time

The codes for the publicly available data should take no more than 20 minutes to run.

The codes for the confidential tax data (on the server) should run in about 8 hours. To achieve this speed, we made a simplifying tweak in the synthetic difference-in-difference estimation (SDID) [setting a low number of bootstrap iterations] which implies that the present version of the code will replicate the point estimates for the SDID estimations (and everything else in the paper and appendix) but will not correctly replicate the SDID standard errors. We suggest that the replicator first runs this version of the code, to make sure everything executes, and then reruns it to obtain the right standard errors for the SDID.

To obtain the same SDID standard errors as in the paper and appendix, change the globals "run_time" and "replication" in 0_master_replication.do to 50 and 500 respectively (note that each of the globals appears multiple times in this code - lines 163,164,195,196,216, and 217). After this change, the code will take 3-4 days to run.

### List of Exhibits

Given the number and complexity of the exhibits we provide the following index containing the specific filename as produced by the code for each figure/table. Additionally, in the ADMIN_DATA_CODES folder we included a .tex file named "Compiler_admin_data_exhibits" which can be executed in the server to create a PDF containing all exhibits from the admin data in the same order and numbering as in the paper. Please copy the compiler file into the output folder "D:_AEJ and execute it once the code is done running.

Figures and Tables Index:

Main Paper: Figure 1. AI. trend_log_countn_t_n02016 AII. trend_log_tot_sales_t_n02016 AIII.trend_log_count_pos_n02016 AIV. trend_log_count_firms_n02016 BI. trend_d_countn_t_d1_n02016. BII. trend_d_tot_sales_t_d1_n02016 BIII.trend_d_count_pos_d1_n02016. BIV. trend_d_count_firms_d1_n02016

Figure 2. AI. RDrobust_log_count_trans_80_deseason_bc2_2015_main AII. RDrobust_log_tot_sales_80_deseason_bc2_2015_main AIII.RDrobust_log_count_pos_total_80_deseason_bc2_2015_main AIV. RDrobust_log_count_firms_total_80_deseason_bc2_2015_main BI. hist_1dtdif_countn_all_random_main BII. hist_1dtdif_tot_sales_all_random_main BIII.hist_1dtdif_count_pos_all_random_main BIV. hist_1dtdif_count_firms_all_random_main

Figure 3. (panel CI requires manual edit described in the readme) AI. did_2lines_s_ta_ab_10_11_log_modif_99_r7_as1_esp10_wholesale AII. did_es_s_ta_ab_10_11_log_modif_99_r7_as1_esp10_wholesale BI. did_2lines_nout_ab_10_11_log_modif_99_r7_as1_esp10_wholesale BII. did_es_nout_ab_10_11_log_modif_99_r7_as1_esp10_wholesale CI. did_2lines_nliab_ab_10_11_log_modif_99_r7_as1_esp10_wholesale_edit CII. event_sdid_wholesale_log_modif_nliab_10_ab_r100_99_r7_10

Figure 4. A. subsidy_firms_share B. event_raw_b_c_2 C. scardsales_density_ms_3_line

Table 1. did_robust_table_logmodif_mixed

Appendix A.

Table A1. Non-data generated

Table A2. Non-data generated

Table A3 sumstat_uruguay_f3_ub_r8

Figure A1 (public data code) AI. account_2011_replication_4 AII. findex_3_4 B. log_informal_2011

Figure A2 Non-data generated

Figure A3. A. ui_density_full_2014 (left) and ui_density_2014 (right) B. ui_density_full_2015 (left) and ui_density_2015 (right)

Figure A4. Non-data generated

Figure A5. Non-data generated

Figure A6. Non-data generated

Figure A7. A. did_vat_rebate_2014 B. did_count1_th_2014 C. did_tot_venta1_2014 D. did_rebate_sale_2_2014

Figure A8. Non-data generated

Figure A9. (public data code) A. findex_1_4_q B. findex_3_4_q

Figure A10 vat_filers_month

Appendix B.

Figure B1. AI. trend_log_countn_n02016 AII.trend_log_tot_sales_n02016. BI. trend_log_countn_t_n02016 BII. trend_log_tot_sales_t_n02016

Figure B2. A. RDrobust_firm_log_count_trans_80_1_all B. RDrobust_firm_log_tot_sales_80_1_all

Figure B3. (his produce all outcomes) AI. RDrobust_log_count_pos_total_80_deseason_bc1_2015_low AII.

RDrobust_log_count_firms_total_80_deseason_bc1_2015_low BI.
hist_1dtdif_count_pos_all_random_main_low BII.
hist_1dtdif_count_firms_all_random_main_low

Figure B4. AI. robust_log_count_trans_1_d_w_0815 AII. robust_log_count_trans_2_d_w_0815
AIII. robust_log_count_trans_3_d_w_0815 BI. robust_log_tot_sales_1_d_w_0815 BII.
robust_log_tot_sales_2_d_w_0815 BIII. robust_log_tot_sales_3_d_w_0815 CI.
robust_log_count_pos_total_1_d_w_0815 CII. robust_log_count_pos_total_1_d_w_0815 CIII.
robust_log_count_pos_total_1_d_w_0815

Figure B5. AI. robust_log_count_trans_1_d_w AII. robust_log_count_trans_2_d_w AIII.
robust_log_count_trans_3_d_w BI. robust_log_tot_sales_1_d_w BII.
robust_log_tot_sales_2_d_w BIII. robust_log_tot_sales_3_d_w CI.
robust_log_count_pos_total_1_d_w CII. robust_log_count_pos_total_1_d_w CIII.
robust_log_count_pos_total_1_d_w

Figure B6. AI. RDrobust_log_count_trans_80_deseason_bc2_2015_jan13 AII.
RDrobust_log_tot_sales_80_deseason_bc2_2015_jan13
AIII.RDrobust_log_count_pos_total_80_deseason_bc2_2015_jan13 AIV.
RDrobust_log_count_firms_total_80_deseason_bc2_2015_jan13 BI.
hist_1dtdif_countn_all_random_jan13 BII. hist_1dtdif_tot_sales_all_random_jan13
BIII.hist_1dtdif_count_pos_all_random_jan13 BIV.
hist_1dtdif_count_firms_all_random_jan13

Figure B7.
AI. placebo_rdd_lntrans_1 AII. placebo_rdd_lnsales_1 AIII.placebo_rdd_lnpos_1 AIV.
placebo_rdd_lnfirms_1 BI. placebo_rdd_lntrans_2 BII. placebo_rdd_lnsales_2
BIII.placebo_rdd_lnpos_2 BIV.placebo_rdd_lnfirms_2

Figure B8. (Public Data Code) Missing A. argentina_logtrans B. argentina_trans_d

Figure B9. A. hist_1dtdif_log_count_transd_70 B. hist_1dtdif_log_tot_salesd_70 C.
hist_1dtdif_log_count_pos_totald_70 D. hist_1dtdif_log_count_firms_totald_70

Figure B10. AI. trend_log_rtaxsale_n02016 AII. trend_log_rnetliab_n02016 BI.
trend_d_rtaxsale_d1_n02016 BII. trend_d_rnetliab_d1_n02016 CI.
RDrobust_m_log_rtaxsale_16_deseason_bc2 CII. RDrobust_m_log_rnetliab_16_deseason_bc2

Table B1. rd_running_aggregation

Table B2. RD_cutweeks_d_short

Table B3. RD_steps_comparisson

Table B4. rdd_autoregression_lag1

Table B5. rdd_autoregression_lag2

Table B6. rdd_autoregression_prais

Figure B11. A. hist_rebate_fraction_month9_122 B. bar_rebates_9 C. hist_rebate_fraction_9

Table C8. triple_did_stata_n_ab_rd_final

Appendix D.

Figure D1. A. count_transactions_2013 B. tot_venta_2013 C. count_pos_2013

Figure D2. AI. hascard_2012_share_T_2009_b AII. hascard_2012_share_T_2009_ub BI. hascard_2012_share_rw_2009_b BII. hascard_2012_share_rw_2009_ub

Figure D3 A. did_month_count_trans_b_log_modif_jan2012 B. did_month_tot_sales_b_log_modif_jan2012 C. did_month_hascard_b_log_modif_jan2012

Figure D4. costarica_cardsales_congraph_replication

Figure D5. (Public Data Code) A. conditional_simple B. formality_share C. debit_card_exp_share

Figure D6. AI. cost_rebates_nominal AII. cost_rebates_total AIII. cost_rebates_firms BI. cost_subsidy_nominal BII. cost_subsidy_total BIII. cost_subsidy_firms

Appendix E. Table E1. differences_hascard

Table E2. differences_hascard_retail

Table E3. hazard_hascard

Table E4. a. FirmFE_hascard b. FirmFE_hascard2

As mentioned in the readme, the simplified version of the code does not reproduced exactly the SE synthetic difference-in-difference estimation (SDID). The following exhibits are impacted by the simplification.
–> Fig 3, Panel CII. –> Table 1, Panel C –> Figure B13, panel C –> Figure B14, panel C –> Table C1, columns 9,10,11,12 –> Table C2, columns 9,10,11 –> Table C3, columns 9,10,11 –> Figure C1, panel CI and CII –> Table C6, columns 7,8,9 –> Figure C2, panel CII, –> Table C7, columns 9,10,11 –> Figure C6, all panels.