VAT refunds and firms' performance: Evidence from a withholding reform in Honduras

Code Repository Documentation

David Pineda Pinto
Utah State University
david.pineda@usu.edu

Jose Carlo Bermúdez
PUC-Chile & DIME, World Bank
jcbermudez@uc.cl

Thiago Scot*
DIME, World Bank
tscot@worldbank.org

*Corresponding author.

Contents

1 Overview

The codes in this repository replicate the tables and figures from "VAT refunds and firms' performance: Evidence from a withholding reform in Honduras", by Pineda Pinto, Bermúdez, and Scot. The replication folder contains the codes to go from the raw (anonymized) 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 provides details on 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 the 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 Honduran tax authority, Servicio de Administración de Rentas (SAR), owns the administrative data for this project. The data were made available exclusively for this research project through the corresponding author and government agency collaboration agreements.

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 Thiago Scot (tscot@worldbank.org). We will provide reasonable assistance to requests for clarification and replication. Researchers interested in obtaining the data for their own analyses can directly contact SAR.

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 Honduran Tax Authority (SAR). Raw data is available mainly from 2011 to 2019 and includes:

- **Monthly withheld taxes.** This includes information coming from an informative monthly tax form (DMR by its acronyms in Spanish for *Declaración Mensual de Retenciones*), including Value Added Tax (VAT) withholding made by Credit and Debit Card (DCC) providers, the government, and large taxpayers.
- **Monthly VAT records.** This includes the universe of formal firms in Honduras filling and submitting a tax form with VAT debits, credits, tax liabilities, and unrefunded balance.
- **Annual Corporate Income Tax (CIT) records.** Dataset on yearly corporate income tax submitted by firms, including balance sheet, income, expenses, tax liabilities, and tax credits and exemptions.
- **Annual Personal Income Tax (PIT) records.** Dataset on yearly corporate income tax submitted by individuals, and small businesses (single-owned) including balance sheet, income, expenses, tax liabilities, and tax credits and exemptions.
- **Tax revenue data.** Dataset with monthly VAT revenue.
- **Risk score data.** Dataset on risk compliance score for the universe of Honduran taxpayers. Risk is assessed using SAR's internal risk model (*Modelo de Gestión de Riesgo de Honduras*, MGR-H) that considers both discrepancies between declared income by taxpayers and information reported by third-parties, as well as anomalies, defined as outcomes that seem inconsistent with other similar tax units.
- **Data from TADAT.** Data on VAT refund payments. Manually gathered from Tax Administration Diagnostic Assessment Tool (TADAT) reports (see TADAT web page).
- **B-Ready reports.** Data on VAT delays. Manually gathered from Enterprise Surveys web page.

3 Datasets

This section describes every dataset used for replicating the results. Table 1 displays the set of raw datasets (found in `Rawdata` 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 `Inputs` folder of the replication package). Those files that are non-available for public disclosure are flagged with a "No" in the last column, while publicly available datasets provided in the replication repository are flagged with a "Yes".

Table 1: Raw datasets

Dataset	**Description**	**Notes**	**Provided**
Base_IM_ATC_CGIT_IT_AE1.dta	Dataset on yearly CIT.	Confidential	No
Base_Mensual_ISV_201101_201912.dta	Dataset with monthly VAT forms.	Confidential	No
Base_Mensual_SER_201101_201912.dta	Dataset with monthly payments from different sources of withholdings.	Confidential	No
base_TADAT.xlsx	Dataset including refunds not paid/declined within 30 days, and VAT refunds as % of VAT gross revenue. Only includes countries that have been subject to a TADAT assessment.	Public. Gathered from TADAT web page	Yes
Bready.xlsx	Dataset including	Public. Gathered from B-Ready.	Yes
ISRPN11_20_initialclean.dta	Dataset on yearly PIT.	Confidential	No
Recaudación_ISV_2016-2019.dta	Dataset with monthly VAT revenue.	Confidential	No
risk_score.dta	Dataset on risk compliance score for the universe of Honduran taxpayers.	Confidential	No

Table 2: Processed datasets

Dataset	Description	Notes	Provided
Firm_traits.dta	Processed dataset including relevant characteristics of firms like economic activity and risk score.	Confidential	No
ISV_panel.dta	Clean dataset at the taxpayer level only including outcomes of interest from VAT records.	Confidential	No
TaxCollection_2016_2019.dta	Clean dataset at the taxpayer level only including taxpayers that have a report on VAT collection.	Confidential	No
ISV_panel_vars.dta	Processed dataset at the taxpayer-month level including relevant outcomes created from VAT records, ready for analysis.	Confidential	No
ISRPJ_workingdatabase.dta	Processed dataset at the taxpayer-year level including relevant outcomes created from CIT records, ready for analysis.	Confidential	No
ISRPN_workingdatabase.dta	Processed dataset at the taxpayer-year level including relevant outcomes created from PIT records, ready for analysis.	Confidential	No
panel_definitions_database.dta	Processed dataset creating variables identifying different sample balancedness at the monthly and quarterly level.	Confidential	No
firmlevel_TreatStatus.dta	Processed dataset including the treatment/control status of every firm and across different sample balancedness, according to four different criteria of DCC usage and/or withholding usage.	Confidential	No
ISV_reform_panel.dta	Processed dataset combining VAT records, sample balancedness indicators, and treatment/control status. Ready for the empirical analysis of the reform.	Confidential	No
shares_quarterly_2016sales.dta	Processed dataset including shares of sales per quarter in 2016.	Confidential	No
shares_monthly_2016sales.dta	Processed dataset including shares of sales per month in 2016.	Confidential	No
Annual_sales.dta	Proccesed dataset including income quartiles for the pre-reform period and different treatment/control groups. At this point, we define final income as the max between yearly VAT turnover and total turnover reported in income tax records.	Confidential	No
heterogeneity_traits.dta	Processed dataset identifying cash-constrained and capital-intensive firms. Only for baseline sample.	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 `Stata` version. 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 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. With those settings, a replicator could expect the whole code to run in 83 minutes. However, replicating the empirical analysis only takes 61 minutes.

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_rawdata`

`2_codes`

`3_analysis`

- `1_inputs (working datasets)`
- `2_figures`
- `3_tables`

We provide the do-files to create the datasets used in the project and the subsequent analysis (from `1_rawdata` and after processing are stored in `1_inputs (working datasets)`). In the master do-file, we set the `cd path` for every so the above folder tree can run easily. The codes described in section 7 rely on this folder structure.

5.2 Before replication

Before running the code and to make the replication process more efficient, in line 54 of section "`Set working paths`" in the master do-file the reviewer must insert the user name and the location where raw data and output files are located. Suppose the replicator gets access to our de-identified data. In that case, we'll share a carpet called `Honduras_project_replication`, including the raw datasets needed for replication, but also, including empty sub-carpets for outputs, with the same names as displayed throughout the code. In addition, the reviewer must insert the path where do files are located (this is not necessarily the same location as the working path if do files are directly retrieved from the GitHub repository of the project).

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	`table_shares_top.tex`	`3a_Descriptives_ISV.do`
Table 2	`summary.tex`	`5c_Descriptives_Reform_Table.do`
Table 3	`first_stage_prepost1.tex`	`6a_Analysis_Reform_ISV.do`
Table 4	`second_stage_prepost1.tex`	`6a_Analysis_Reform_ISV.do`
Table 5	`investment_prepost1.tex`	`6a_Analysis_Reform_ISR.do`
Table 6	`performance_prepost1.tex`	`6a_Analysis_Reform_ISR.do`
Table A1	`table_descriptive_firmyear_fullperiod.tex`	`3a_Descriptives_ISV.do`
Table A2	`vat_robustness1.tex`	`6b_Robustness_ISV.do`
Table A3	`performance_robustness1.tex`	`6b_Robustness_ISR.do`
Table A4	`vat_robustness3.tex`	`6b_Robustness_ISV.do`
Table A5	`investment_robustness3.tex`	`6b_Robustness_ISR.do`
Table A6	`performance_robustness3.tex`	`6b_Robustness_ISR.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	**Panels**	**Output filename**	**Code**
Figure 1	a-d	`refunds_in_time.pdf`, `refunds_share.pdf`, `refunds_weeks.pdf`, `refunds_percent_firms_complicated.pdf`	`3b_Descriptives_CrossCountry_Refunds.do`
Figure 2		`accumulated_balance.pdf`	`3a_Descriptives_ISV.do`
Figure 3	a,b	`change_share_withheld_FULL.pdf`, `change_retention_all_FULL.pdf`	`5a_Descriptives_Reform_Figures.do`
Figure 4	a-d	`mean1_withholding.pdf`, `did1_withholding.pdf`, `mean1_etr.pdf`, `did1_etr.pdf`, `mean1_dummy_unrefunded.pdf`, `did1_dummy_unrefunded.pdf`, `mean1_unref_balance.pdf`, `did1_unref_balance.pdf`	`6a_Analysis_Reform_ISV.do`
Figure 5	a-c	`mean1_tax_sales.pdf`, `did1_tax_sales.pdf`, `mean1_tax_purchases.pdf`, `did1_tax_purchases.pdf`, `mean1_value_added.pdf`, `did1_value_added.pdf`	`6a_Analysis_Reform_ISV.do`
Figure 6	a-c	`mean1_inv_net2.pdf`, `did1_inv_net2.pdf`, `mean1_prob_inv_net.pdf`, `did1_prob_inv_net.pdf`, `mean1_ppe_net2.pdf`, `did1_ppe_net2.pdf`	`6a_Analysis_Reform_ISR.do`
Figure 7	a-c	`mean1_cash_stock.pdf`, `did1_cash_stock.pdf`, `mean1_labor_costs.pdf`, `did1_labor_costs.pdf`, `mean1_pre_tax_profits.pdf`, `did1_pre_tax_profits.pdf`	`6a_Analysis_Reform_ISR.do`
Figure 8		`het1_etr.pdf`	`6d_Heterogeneity_ISV.do`
Figure 9	a-c	`het1_inv_net2.pdf`, `het1_prob_inv_net.pdf`, `het1_ppe_net2.pdf`	`6d_Heterogeneity_ISR.do`
Figure A1	a,b	`composition_sales.pdf`, `composition_purchases.pdf`	`3a_Descriptives_ISV.do`
Figure A2		`coefplot_reg_determinants2.pdf`	`3a_Descriptives_ISV.do`
Figure A3		`heraldo.pdf`, `prensa.pdf`, `tiempo.pdf`	Collected from media websites
Figure A4	a	`withhold_amounts_compare.pdf`	`5a_Descriptives_Reform_Figures.do`
Figure A4	b	`withholding_isv_ser.pdf`	`5b_Descriptives_SERaggregates.do`
Figure A5		`cdf_share_withheld_fullperiod.pdf`	`5a_Descriptives_Reform_Figures.do`
Figure A6		`change_share_withheld_all.pdf`	`5a_Descriptives_Reform_Figures.do`
Figure A7	a,b	`firms_entry_atc.pdf`, `firms_claim_atc.pdf`	`5a_Descriptives_Reform_Figures.do`
Figure A8	a,b	`density_with.pdf`, `density_with_usage.pdf`	`5a_Descriptives_Reform_Figures.do`
Figure A9		`histo_sharewith_tc.pdf`	`5a_Descriptives_Reform_Figures.do`
Figure A10	a-c	`vat_change.pdf`, `vat_change_firms_withheld.pdf`, `vat_change_firms_withheld_normalized.pdf`	`5a_Descriptives_Reform_Figures.do`
Figure A11	a,b	`mean1_withholding_cards.pdf`, `did1_withholding_cards.pdf`	`6a_Analysis_Reform_ISV.do`

Exhibit Name	Panels	Output filename	Code
Figure A12	a-c	mean1_revenue.pdf, did1_revenue.pdf, mean1_purchases.pdf, did1_purchases.pdf, mean1_exem_purchases.pdf, did1_exem_purchases.pdf	6a_Analysis_Reform_ISV.do
Figure A13	a-c	mean1_inv_gross2.pdf, did1_inv_gross2.pdf, mean1_prob_inv_net.pdf, did1_prob_inv_net.pdf, mean1_ppe_net2.pdf, did1_ppe_net2.pdf	6a_Analysis_Reform_ISR.do
Figure A14	a-d	mean1_current_assets.pdf, did1_current_assets.pdf, mean1_noncurrent_assets.pdf, did1_noncurrent_assets.pdf, mean1_current_liab.pdf, did1_current_liab.pdf, mean1_noncurrent_liab.pdf, did1_noncurrent_liab.pdf	6a_Analysis_Reform_ISR.do
Figure A15		intensity.pdf	5a_Descriptives_Reform_Figures.do
Figure A16	a,b	var_dcc_usage_control.pdf, var_dcc_usage_treatment.pdf	5a_Descriptives_Reform_Figures.do
Figure A17	a-d	did3_withholding.pdf, did3_etr.pdf, did3_dummy_unrefunded.pdf, did3_unref_balance.pdf	6c_Robustness_ISV_Treat.do
Figure A18	a-e	did3_inv_net2.pdf, did3_prob_inv_net.pdf, did3_cash_stock.pdf, did3_labor_costs.pdf, did3_after_tax_profits.pdf	6c_Robustness_ISR_Treat.do
Figure A19	a-d	het1_withholding.pdf, het1_tax_total.pdf, het1_dummy_unrefunded.pdf, het1_unref_balance.pdf	6d_Heterogeneity_ISV.do
Figure A20	a-c	het1_tax_sales.pdf, het1_tax_purchases.pdf, het1_value_added.pdf	6d_Heterogeneity_ISV.do
Figure A21	a-d	het1_cash_stock.pdf, het1_labor_costs.pdf, het1_pre_tax_profits.pdf, het1_after_tax_profits.pdf	6d_Heterogeneity_ISR.do
Figure A22	a-c	het1_inv_gross2.pdf, het1_prob_inv_gross.pdf, het1_ppe_gross2.pdf	6d_Heterogeneity_ISR.do
Figure A23	a,b	liquidity_histogram_pooled.pdf, liquidity_histogram_split.pdf	4c_PrepareTime_Invariant.do
Figure A24	a,b	capitalint_histogram_pooled.pdf, capitalint_histogram_split.pdf	4c_PrepareTime_Invariant.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 adjust the links in these files centrally without needing to adjust 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

- `1a_Clean_Databases.do`

  This do-file takes raw de-identified datasets from DCC payments, VAT forms, withholdings, and tax revenue, harmonizes them, and creates intermediate datasets that we use later for further analysis.

  Inputs:

  $Rawdata/Base_Mensual_ISV_201101_201912.dta
  $Rawdata/risk_score.dta
  $Rawdata/Base_Mensual_ISV_201101_201912.dta
  $Rawdata/Base_Mensual_SER_201101_201912.dta
  $Rawdata/Recaudación_ISV_20162019.dta

  Outputs:

  $Inputs/Firm_traits.dta
  $Inputs/ISV_panel.dta
  $Inputs/Base_Mensual_SER_201101_201912.dta
  $Inputs/TaxCollection_2016_2019.dta

- `1b_Create_ISV_panelvars.do`

  This do-file takes processed de-identified VAT records, harmonizes them, and creates an intermediate dataset for further analysis.

  Input:

  $Inputs/ISV_panel.dta

  Output:

  $Inputs/ISV_panel_vars.dta

- `1c_Clean_ISRPJ.do`

  This do-file raw de-identified dataset from corporate income tax data, clean, harmonize, and build balance sheet outcomes. The output is ready for analysis.

  Input:

  $Rawdata/Base_IM_ATC_CGIT_IT_AE1.dta

Output:

$Inputs/ISRPJ_workingdatabase.dta

- 1d_Clean_ISRPN.do

  This do-file raw de-identified dataset from personal income tax data, clean, harmonize, and build balance sheet outcomes. The output is ready for analysis.

  Input:

  $Rawdata/ISRPN11_20_initialclean.dta

  Output:

  $Inputs/ISRPN_workingdatabase.dta

- 4a_PrepareBalancedDefinition.do

  This do-file creates identified samples according to different levels of balancedness and periods.

  Input:

  $Inputs/ISV_panel_vars.dta
  $Inputs/ISRPN_workingdatabase.dta
  $Inputs/ISRPJ_workingdatabase.dta

  Output:

  $Inputs/panel_definitions_database.dta

- 4b_PreparePanelReform_treat.do

  This do-file identifies different treatment and control groups.

  Input:

  $Inputs/ISV_panel_vars.dta
  $Inputs/panel_definitions_database.dta

  Output:

  $Inputs/firmlevel_TreatStatus.dta

- 4c_PrepareTime_Invariant.do

  This do-file creates true income for firms in the sample, and additional time-invariant categories such as pre-reform quartile of income, liquidity constraints, and capital intensity.

  Input:

  $Inputs/ISV_panel_vars.dta
  $Inputs/firmlevel_TreatStatus.dta
  $Inputs/Firm_traits.dta
  $Inputs/ ISRPN_workingdatabase.dta
  $Inputs/ISRPJ_workingdatabase.dta

  Output:

  $Inputs/Annual_sales.dta

$Inputs/heterogeneity_traits.dta
$Figures/liquidity_histogram_pooled.pdf
$Figures/liquidity_histogram_split.pdf
$Figures/capitalint_histogram_pooled.pdf
$Figures/capitalint_histogram_split.pdf

- 4d_Prepare_Sales_Shares.do
  This do-file creates variables with the share of monthly/quarterly sales in 2016, the year before the reform. These shares are then used for normalization of outcomes.

  Input:
  $Inputs/ISV_panel_vars.dta

  Output:
  $Inputs/shares_quarter_2016sales.dta
  $Inputs/shares_month_2016sales.dta

- 4e_PreparePanelReform_ISV.do
  This do-file creates dataset from VAT records ready for analysis.

  Input:
  $Inputs/ISV_panel_vars.dta
  $Inputs/firmlevel_TreatStatus.dta
  $Inputs/shares_quarter_2016sales.dta
  $Inputs/Annual_sales.dta

  Output:
  $Inputs/ISV_reform_panel_quarter.dta

- 4e_PreparePanelReform_ISR.do
  This do-file creates dataset from income tax records ready for analysis.

  Input:
  $Inputs/ISV_panel_vars.dta
  $Inputs/ISRPN_workingdatabase.dta
  $Inputs/ISRPN_workingdatabase.dta
  $Inputs/firmlevel_TreatStatus.dta
  $Inputs/Annual_sales.dta
  $Inputs/ heterogeneity_traits.dta

  Output:
  $Inputs/ISR_reform_panel.dta

7.2 Analysis

- 3a_Descriptives_ISV.do

  This do-file uses processed de-identified VAT records to produce tables 1 and A1 and figures 2, A1, and A2.

  Input:

  $Inputs/ISV_reform_panel.dta
  $Inputs/ISRPN_workingdatabase.dta
  $Inputs/ISRPJ_workingdatabase.dta
  $Inputs/Firm_traits.dta

  Output:

  $Tables/table_shares_top.tex
  $Tables/table_descriptive_firmyear_fullperiod.tex
  $Figures/composition_sales.pdf
  $Figures/composition_purchases.pdf
  $Figures/accumulated_balance.pdf
  $Figures/histo_share_slides1.pdf
  $Figures/scatter_share_intq.pdf
  $Figures/box_share_sectors.pdf
  $Figures/coefplot_reg_determinants2.pdf

- 3b_Descriptives_CrossCountry_Refunds.do

  This do-file produces Figure 1. The database was constructed manually using TADAT and B-Ready country reports.

  Input:

  $Inputs/base_TADAT.xlsx
  $Inputs/Bready.xlsx

  Output:

  $Figures/refunds_in_time.pdf
  $Figures/refunds_share.pdf
  $Figures/refunds_weeks.pdf
  $Figures/refunds_percent_firms_complicated.pdf

- 5a_Descriptives_Reform_Figures.do

  This do file produces figures 3, A4, A5, A6, A7, A8, A9, A10, A15, and A16.

  Input:

  $Inputs/ISV_panel_vars.dta
  $Inputs/firmlevel_TreatStatus.dta

  Output:

  $Figures/change_share_withheld_FULL.png
  $Figures/change_retention_all_FULL.png
  $Figures/withhold_amounts_compare.png
  $Figures/cdf_share_withheld_fullperiod.png

$Figures/change_share_withheld_all.png
$Figures/density_with.pdf
$Figures/density_with_usage.pdf
$Figures/histo_sharewith_tc.pdf
$Figures/share_claiming.png
$Figures/firms_entry_atc.pdf
$Figures/firms_claim_atc.pdf
$Figures/var_dcc_usage_control.pdf
$Figures/var_dcc_usage_treatment.pdf
$Figures/intensity.pdf
$Figures/vat_change.pdf
$Figures/vat_change_firms_withheld.pdf
$Figures/vat_change_firms_withheld_noscale.pdf
$Figures/vat_change_firms_withheld_normalized.pdf

- 5b_Descriptives_SERaggregates.do
  This do-file produces Figure A4b.

  Input:
  $Inputs/Base_Mensual_SER_201101_201912.dta
  $Inputs/ISV_panel_vars.dta

  Output:
  $Figures/withholding_isv_ser.pdf

- 5c_Descriptives_Reform_Table.do
  This do file produces Table 3 with summary statistics in 2016 for the baseline sample included in the empirical analysis of the reform.

  Input:
  $Inputs/ISV_reform_panel.dta
  $Inputs/firmlevel_TreatStatus.dta
  $Inputs/ISRPN_workingdatabase.dta
  $Inputs/ISRPJ_workingdatabase.dta
  $Inputs/Annual_sales.dta
  $Inputs/Firm_traits.dta

  Output:
  $Tables/summary.tex

- 6a_Analysis_Reform_ISV.do
  This do-file runs DiD models and produces Figures 4, 5, A11, and A12. Also produces Tables 3 and 4.

  Input:
  $Inputs/ISV_reform_panel_quarter.dta

$Inputs/firmlevel_TreatStatus.dta
$Inputs/Annual_sales.dta
$Inputs/heterogeneity_traits.dta

Output:

$Tables/first_stage_prepost1.tex
$Tables/second_stage_prepost1.tex
$Figures/mean1_value_added.pdf
$Figures/mean1_tax_total.pdf
$Figures/mean1_etr.pdf
$Figures/mean1_dummy_unrefunded.pdf
$Figures/mean1_unref_balance.pdf
$Figures/mean1_revenue.pdf
$Figures/mean1_tax_sales.pdf
$Figures/mean1_purchases.pdf
$Figures/mean1_tax_purchases.pdf
$Figures/mean1_exem_purchases.pdf
$Figures/mean1_withholding_cards.pdf
$Figures/mean1_withholding.pdf
$Figures/did1_value_added.pdf
$Figures/did1_tax_total.pdf
$Figures/did1_etr.pdf
$Figures/did1_dummy_unrefunded.pdf
$Figures/did1_unref_balance.pdf
$Figures/did1_revenue.pdf
$Figures/did1_tax_sales.pdf
$Figures/did1_purchases.pdf
$Figures/did1_tax_purchases.pdf
$Figures/did1_exem_purchases.pdf
$Figures/ did1_withholding_cards.pdf
$Figures/did1_withholding.pdf

- 6a_Analysis_Reform_ISR.do

  This do file runs DiD models and produces figures 6, 7, A13, and A14. Also produces tables 5 and 6.

  Input:

  $Inputs/ISR_reform_panel.dta
  $Inputs/firmlevel_TreatStatus.dta
  $Inputs/heterogeneity_traits.dta
  $Inputs/Annual_sales.dta

  Output:

  $Tables/balance_prepost1.tex
  $Tables/investment_prepost1.tex
  $Tables/performance_prepost1.tex

$Figures/mean1_after_tax_profits.pdf
$Figures/mean1_pre_tax_profits.pdf
$Figures/mean1_labor_costs.pdf
$Figures/mean1_cash_stock.pdf
$Figures/mean1_prob_inv_net.pdf
$Figures/mean1_prob_inv_gross.pdf
$Figures/mean1_inv_net2.pdf
$Figures/mean1_inv_gross2.pdf
$Figures/mean1_ppe_net2.pdf
$Figures/mean1_ppe_gross2.pdf
$Figures/mean1_noncurrent_liab.pdf
$Figures/mean1_current_liab.pdf
$Figures/mean1_current_assets.pdf
$Figures/mean1_noncurrent_assets.pdf
$Figures/mean1_deductions.pdf
$Figures/did1_after_tax_profits.pdf
$Figures/did1_pre_tax_profits.pdf
$Figures/did1_labor_costs.pdf
$Figures/did1_cash_flow.pdf
$Figures/did1_prob_inv_net.pdf
$Figures/did1_prob_inv_gross.pdf
$Figures/did1_inv_net2.pdf
$Figures/did1_inv_gross2.pdf
$Figures/did1_ppe_net2.pdf
$Figures/did1_ppe_gross2.pdf
$Figures/did1_noncurrent_liab.pdf
$Figures/did1_current_liab.pdf
$Figures/did1_noncurrent_assets.pdf
$Figures/did1_current_assets.pdf
$Figures/did1_deductions.pdf

- 6b_Robustness_ISV.do
  Conducts robustness checks running regression models on different samples and alternative winsorizing using VAT outcomes. Produces table A2 and A4.

  Input:
  $Inputs/ISV_reform_panel.dta
  $Inputs/Annual_sales.dta
  $Inputs/ shares_quarterly_2016sales.dta

  Output:
  $Tables/vat_robustness1.tex
  $Tables/vat_robustness3.tex

- 6b_Robustness_ISR.do

Conducts robustness checks running regression models on different samples and alternative winsorizing using balance sheets. Produces tables A3, A5, and A6.

Input:

$Inputs/ISV_reform_panel.dta
$Inputs/ISRPN_workingdatabase.dta
$Inputs/ISRPJ_workingdatabase.dta
$Inputs/Annual_sales.dta

Output:

$Tables/performance_robustness1.tex
$Tables/investment_robustness3.tex
$Tables/performance_robustness3.tex

- 6c_Robustness_ISV_Treat.do

  Conducts robustness checks for VAT outcomes changing the baseline definition of treatment and control groups. Produces figure A17.

  Input:

  $Inputs/ISV_reform_panel_quarter.dta
  $Inputs/firmlevel_TreatStatus.dta
  $Inputs/heterogeneity_traits.dta
  $Inputs/Annual_sales.dta

  Output:

  $Figures/did3_value_added.pdf
  $Figures/did3_tax_total.pdf
  $Figures/did3_etr.pdf
  $Figures/did3_dummy_unrefunded.pdf
  $Figures/did3_unref_balance.pdf
  $Figures/did3_revenue.pdf
  $Figures/did3_tax_sales.pdf
  $Figures/did3_purchases.pdf
  $Figures/did3_tax_purchases.pdf
  $Figures/did3_exem_purchases.pdf
  $Figures/did3_withholding_cards.pdf
  $Figures/did3_withholding.pdf

- 6c_Robustness_ISR_Treat.do

  Conducts robustness checks for balance sheet outcomes changing the baseline definition of treatment and control groups. Produces figure A18.

  Input:

  $Inputs/ISR_reform_panel.dta
  $Inputs/firmlevel_TreatStatus.dta
  $Inputs/heterogeneity_traits.dta
  $Inputs/Annual_sales.dta

Output:

$Figures/did3_after_tax_profits.pdf
$Figures/did3_pre_tax_profits.pdf
$Figures/did3_labor_costs.pdf
$Figures/did3_cash_flow.pdf
$Figures/did3_prob_inv_net.pdf
$Figures/did3_prob_inv_gross.pdf
$Figures/did3_inv_net2.pdf
$Figures/did3_inv_gross2.pdf
$Figures/did3_ppe_net2.pdf
$Figures/did3_ppe_gross2.pdf
$Figures/did3_noncurrent_liab.pdf
$Figures/did3_current_liab.pdf
$Figures/did3_noncurrent_assets.pdf
$Figures/did3_current_assets.pdf
$Figures/did3_deductions.pdf

- 6d_Heterogeneity_ISV.do

  Run conditional DiD for heterogeneity analysis. Produces figures 8, A19, and A20.

  Input:

  $Inputs/ISV_reform_panel_quarter.dta
  $Inputs/firmlevel_TreatStatus.dta
  $Inputs/heterogeneity_traits.dta
  $Inputs/Annual_sales.dta

  Output:

  $Figures/het1_value_added.pdf
  $Figures/het1_tax_total.pdf
  $Figures/het1_etr.pdf
  $Figures/het1_dummy_unrefunded.pdf
  $Figures/het1_unref_balance.pdf
  $Figures/het1_revenue.pdf
  $Figures/het1_tax_sales.pdf
  $Figures/het1_purchases.pdf
  $Figures/het1_tax_purchases.pdf
  $Figures/het1_exem_purchases.pdf
  $Figures/het1_withholding_cards.pdf
  $Figures/het1_withholding.pdf

- 6d_Heterogeneity_ISR.do

  Run conditional DiD for heterogeneity analysis. Produces figures 9, A21, and A22.

  Input:

  $Inputs/ISV_reform_panel.dta

$Inputs/ISRPN_workingdatabase.dta
$Inputs/ISRPJ_workingdatabase.dta
$Inputs/heterogeneity_traits.dta
$Inputs/Annual_sales.dta

Output:

$Figures/het1_after_tax_profits.pdf
$Figures/het1_pre_tax_profits.pdf
$Figures/het1_labor_costs.pdf
$Figures/het1_cash_flow.pdf
$Figures/het1_prob_inv_net.pdf
$Figures/het1_prob_inv_gross.pdf
$Figures/het1_inv_net2.pdf
$Figures/het1_inv_gross2.pdf
$Figures/het1_ppe_net2.pdf
$Figures/het1_ppe_gross2.pdf
$Figures/het1_noncurrent_liab.pdf
$Figures/het1_current_liab.pdf
$Figures/het1_noncurrent_assets.pdf
$Figures/het1_current_assets.pdf
$Figures/het1_deductions.pdf

7.3 Programs

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

- `winsor2`
  David Kantor, 2004. "WINSOR2: Stata module to winsorize data," Statistical Software Components S457765, Boston College Department of Economics.

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

- `reghdfe`
  Correia, Sergio. 2017. "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator" Working Paper.

- `ftools`
  Sergio Correia, and Julian Reif. (2023). `sergiocorreia/ftools`: ftools 2.49.1 (08aug2023) (2.49.1). Zenodo.

- `erepost`
  Jann, B. (2007). erepost: Stata module to repost the estimation results.

- `egenmore`
  Nicholas J. Cox, 2000. "EGENMORE: Stata modules to extend the generate function," Statistical Software Components S386401, Boston College Department of Economics.

- `cdfplot`
  Adrian Mander, 2005. "CDFPLOT: Stata module to plot a cumulative distribution function," Statistical Software Components S456409, Boston College Department of Economics, revised 14 Jul 2008.

- `coefplot`
  Jann, B. (2014). Plotting regression coefficients and other estimates. The Stata Journal 14(4): 708-737.

- `wbopendata`
  Azevedo, J. P. (2020). "WBOPENDATA: Stata module to access World Bank databases."

- `gtools`
  Mauricio Caceres Bravo.

- `cic`
  Keith Kranker, 2019. "CIC: Stata module to implement the Athey and Imbens (2006) Changes-in-Changes model" Statistical Software Components S458656, Boston College Department of Economics.

- `outreg2`
  Roy Wada, 2005. "OUTREG2: Stata module to arrange regression outputs into an illustrative table", Statistical Software Components S456416, Boston College Department of Economics, revised 17 Aug 2014.