

# **The Transparency Tax: Discouraging Firm Ownership via Tax Havens in Ecuador**

Code Repository Documentation

Jakob Brounstein

Pierre Bachas

Alex Bajaña

# Contents

|          |                                                                              |           |
|----------|------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Overview</b>                                                              | <b>3</b>  |
| <b>2</b> | <b>Data Availability and Provenance Statements</b>                           | <b>3</b>  |
| 2.1      | Statement about Rights . . . . .                                             | 3         |
| 2.2      | Summary of Availability . . . . .                                            | 3         |
| 2.3      | Details on each Data Source . . . . .                                        | 4         |
| 2.3.1    | SRI administrative data . . . . .                                            | 4         |
| 2.3.2    | Orbis Data . . . . .                                                         | 5         |
| <b>3</b> | <b>Datasets</b>                                                              | <b>7</b>  |
| 3.1      | SRI administrative data . . . . .                                            | 7         |
| 3.2      | Orbis Data . . . . .                                                         | 12        |
| <b>4</b> | <b>Computational Requirements</b>                                            | <b>15</b> |
| 4.1      | Software requirements . . . . .                                              | 15        |
| 4.2      | Memory and runtime requirements . . . . .                                    | 15        |
| 4.2.1    | SRI administrative data . . . . .                                            | 15        |
| 4.2.2    | Orbis data . . . . .                                                         | 15        |
| <b>5</b> | <b>Instructions to replicators</b>                                           | <b>15</b> |
| 5.1      | Folder structure: SRI administrative data . . . . .                          | 15        |
| 5.2      | Before replication: SRI administrative data . . . . .                        | 16        |
| 5.3      | Folder structure: Orbis data . . . . .                                       | 16        |
| 5.4      | Before replication: Orbis data . . . . .                                     | 17        |
| <b>6</b> | <b>List of tables, and figures</b>                                           | <b>18</b> |
| 6.1      | Mapping of tables and analysis code . . . . .                                | 18        |
| 6.2      | Mapping of figures and analysis code . . . . .                               | 19        |
| <b>7</b> | <b>Description of programs and code</b>                                      | <b>20</b> |
| 7.1      | Cleaning data process . . . . .                                              | 20        |
| 7.2      | Empirical analysis . . . . .                                                 | 22        |
| 7.2.1    | Analysis using SRI data . . . . .                                            | 22        |
| 7.2.2    | Supplementary code: ICIJ list construction (run outside the SRI environment) | 25        |
| 7.2.3    | Exhibit construction . . . . .                                               | 25        |
| 7.3      | Analysis using Orbis data . . . . .                                          | 26        |
| 7.4      | Programs . . . . .                                                           | 28        |

# 1 Overview

The codes in this repository replicate the tables and figures from “The Transparency Tax: Discouraging firm ownership via tax havens in Ecuador”, by Brounstein, Bachas, and Bejaña. 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 7.4).

## 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 tax authority from Ecuador, Servicio de Rentas Internas (SRI), owns the administrative data for this project. The data were made available exclusively for this research project through the corresponding author and government 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 Pierre Bachas ([pbachas@worldbank.org](mailto:pbachas@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 SRI.

**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

### 2.3.1 SRI administrative data

The analysis is based exclusively on de-identified administrative raw data provided by the SRI. All raw source files read in `1_replication_build.do` span fiscal years 2012 to 2019 (confirmed by the author team). Based only on the variables and processing steps present in the code, the raw inputs are:

- **APS (`A_APS_<year>.csv`, 2012–2019).** Firm-level shareholder/ownership records. Key variables read by the code: fiscal year, an anonymized filer ID (`identificacion_informante_anon`, used as `firm_id`), anonymized firm and shareholder IDs, ownership level (`nivel_aps`), percentage and effective percentage of participation, a tax-haven-shareholder flag (`paraíso_fiscal_accionista`), shareholder country/country code, and shareholder ID/entity type. Used to build the firm-level ownership panel and the `aps`, `aps_2014`, `aps_2012_2014` compliance/filing indicators.
- **F101 (`F101_<year>.csv`, 2012–2019).** Firm-level annual tax-return records. Variables read: anonymized firm ID, fiscal year, declared revenue (`total_ingresos_1930`), and a corporate-income-tax-liability variable (`impuesto_renta_causado_3600`, read but not used in the final build). Used to build the `f101`, `active`, and `active_2014` (positive-revenue) filters that define the estimation sample.
- **F102 (`f102_processed_all.dta`).** Only the anonymized firm ID (`id_an`) is read from this source in the code; used solely to build a firm-ID list (`f102_id_list.dta`) referenced later when classifying shareholder types.
- **MID – Anexo de Movimiento Internacional de Divisas (`A_MID_<year>.csv`, 2012–2019).** International currency-transfer records. Variables read: anonymized firm ID, transferred amount (`monto_transferido`), transaction type (`tipo_transaccion`), destination/origin country (`país_super_banco_codigo_país`), and an economic-reason code (`motivo_eco_super_bancos / codigo_motivo_economico`). Collapsed to firm  $\times$  transaction-type  $\times$  country  $\times$  reason-code totals and restricted to firms already present in the APS/F101 sample.
- **RUC crosswalk (`crosswalk_ruc_id_update_16082021.dta`).** Per an in-code comment, this is the RUC (taxpayer registry) database; it is used to identify government entities and oil companies that are excluded from the transparency law and regulated under a different regime.
- **CPI by year (`cpi_year.dta`).** Read and copied into the working data directory with no further code-level documentation; based on the file name, appears to be a yearly price index used elsewhere for deflating monetary variables. *Not confirmed by an in-code comment.*
- **ICIJ data (`icij_all.dta`).** Read and copied into the working data directory; used downstream in `icij.do` for the heterogeneity analysis by prominent/major offshore jurisdictions. Unlike the other sources in this list, this is not raw data supplied directly by the SRI: it is the result of a multi-step matching and anonymization process, carried out prior to and outside of the SRI environment, and later anonymized internally by SRI personnel. The steps were:
  1. *ICIJ download.* Three node files were downloaded from the ICIJ Offshore Leaks Database (<https://offshoreleaks.icij.org/pages/database>), Panama Papers release (2016), on 5 March 2022: `nodes-entities.csv`, `nodes-officers.csv`, and `nodes-intermediaries.csv`, each merged to `nodes-addresses` on `node_id`.

2. *Restriction to Ecuador.* Observations were restricted to those with an Ecuadorian presence and names were manually cleaned, producing `ecuador_icij_officer_names.xlsx` and `ecuador_icij_intermediaries.xlsx` (26 May 2022) and `ecuador_icij_entities_names.xlsx` (4 July 2023).
3. *Matching to RUC.* Using the SRI’s taxpayer registry (`ruc_completo_establecimientos.rds`, downloaded 27 May 2022 from the SRI’s public datasets portal, <https://www.sri.gob.ec/datasets>), each of the three Ecuador-restricted files was matched to a RUC (taxpayer ID) by name: first a deterministic exact-name join, then a fuzzy join (cosine distance for officers/intermediaries; Jaro-Winkler distance, in three increasing thresholds, for entities) on the remaining unmatched names. This step was carried out with the R scripts `Functions.R` (31 May 2022), `01_Match_with_RUC.R` (28 June 2022), and `02_Final_format.R` (28 June 2022) (see Section 7), run on the authors’ personal computers – not inside the SRI environment – in the ICIJ list creation project shared under `SRI_replication_11092026/`.
4. *Consolidation and anonymization.* Only the resulting RUC correspondences from the three matched files were consolidated and handed over to SRI personnel, who applied the SRI’s internal anonymization process and delivered the final `icij_all.dta` from within the SRI environment.

Because it links firm/individual-level RUC identifiers to ICIJ-flagged entities, `icij_all.dta` is confidential and cannot be redistributed, even though the original ICIJ source data are public (see Table 2).

- **Country classifications (`países_mid_aps.dta`, `países_clasificados.csv`).** Country-level classification files (e.g., tax-haven, Caribbean, Latin American groupings) used throughout the build to construct the `paraíso_fiscal`, `caribe`, and `latinoamerica` indicators.

### 2.3.2 Orbis Data

Complementary robustness checks is done using data from Orbis<sup>1</sup>, and included the following data sources:

- **Data on Subsidiaries.** Data including a list of foreign subsidiaries located in Ecuador. First criteria includes all active firms. Then, we select those firms that are subsidiaries with foreign shareholders (ownership  $\geq 0.01\%$ ) from any location in the world. A caveat is that the user through which we got access to Orbis, historical data option is not available. Instead, we get a cross section of 17,379 active subsidiaries which have at least one register of direct ownership, but this number decreases as we clean the data.
- **Data on Direct Ownership.** The way we retrieve data on subsidiaries also give us direct links and shares of ownership. Hence, the same data on subsidiaries includes three columns with information on direct ownership: name of direct owner, location of direct owner, and share of direct ownership. This data is collected by choosing the share of ownership of the initial cross section of more than 17k subsidiaries, by the December of each year, from 2014 to 2024.

---

<sup>1</sup>Orbis is a subscription-based global database managed by Bureau van Dijk/Moodys that aims to provide comprehensive financial and corporate information on millions of public and private companies worldwide, including their ownership structures.

- **Data on *Ultimate Ownership*.** This includes description of global ultimate owners including ID of the owner, location, and revenue. Spans from 2016 to 2024. At the moment of downloading the data on Ecuadorian subsidiaries, some variables identifying ultimate ownership (UO) are also included, such as the name of the UO and the country where it is located. This means that we consider all UOs regardless of their size or location. One important caveat to keep in mind is that, unlike what we do with direct owners, where ownership is identified at the end of a given year, we cannot apply the same approach to UOs. This might imply that the UO we observe in 2014 are the same as 2024, and we would not observe any changes in ownership shares over the years.

### 3 Datasets

#### 3.1 SRI administrative data

This section describes every dataset used for replicating the results using tax records. [Table 1](#) displays the raw SRI source files read by `1_replication_build.do`, and [Table 2](#) gives the provider, access instructions, and license for each of these raw sources. [Table 3](#) lists the main intermediate datasets that this same file produces (saved to the confidential working directory `$datadir`, not included in this package). [Table 4](#) lists the analysis-ready datasets produced by the 12 do-files called from `2_replication_analysis.do`; these ARE included in this replication package, under `replication_exhibits_sourcefiles/`, and are consumed directly by `3_replication_exhibits_construction.do` to produce all final tables and figures (see [section 7](#) for the exact input/output mapping of each do-file). Datasets not available for public disclosure are flagged “No” in the last column; datasets included in this repository are flagged “Yes”.

Table 1: Raw datasets (read by `1_replication_build.do`)

| Dataset                                                                   | Description                                                                                                                                                                                                               | Notes                                  | Provided |
|---------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|----------|
| <code>A_APS_&lt;year&gt;.csv</code> (2012–2019)                           | Firm-level shareholder/ownership annex (APS).                                                                                                                                                                             | Confidential (SRI)                     | No       |
| <code>F101_&lt;year&gt;.csv</code> (2012–2019)                            | Firm-level annual tax-return records (F101).                                                                                                                                                                              | Confidential (SRI)                     | No       |
| <code>A_MID_&lt;year&gt;.csv</code> (2012–2019)                           | International currency-transfer annex (MID).                                                                                                                                                                              | Confidential (SRI)                     | No       |
| <code>f102_processed_all.dta</code>                                       | Source of the F102 firm-ID list.                                                                                                                                                                                          | Confidential (SRI)                     | No       |
| <code>crosswalk_ruc_id_update_16082021.dta</code>                         | RUC (taxpayer registry) crosswalk; identifies excluded government/oil firms.                                                                                                                                              | Confidential (SRI)                     | No       |
| <code>cpi_year.dta</code>                                                 | Yearly price index (not documented in-code).                                                                                                                                                                              | Not redistributed here                 | No       |
| <code>icij_all.dta</code>                                                 | RUC (taxpayer ID) correspondences for ICIJ-flagged entities/officers/intermediaries with an Ecuadorian presence, matched by name and anonymized by SRI personnel (see <a href="#">section 2</a> for the full derivation). | Confidential (derived, SRI-anonymized) | No       |
| <code>países_mid_aps.dta</code> ,<br><code>países_clasificados.csv</code> | Country classifications (tax haven / Caribbean / Latin America).                                                                                                                                                          | Not redistributed here                 | No       |

Table 2: Data source details for the raw datasets in Table 1

| Filename(s)                                                                                                                  | Source                                                                                                      | URL / access instructions                                                                                                                                                                                      | Access date                                                       | Citation                                                                                                                                                                                                             | License                                                                                                                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A_APS_<year>.csv,<br>F101_<year>.csv,<br>A_MID_<year>.csv,<br>f102_processed_all.dta<br>crosswalk_ruc_id_update_16082021.dta | Servicio de Rentas Internas (SRI), Ecuador                                                                  | Not publicly available. Confidential administrative microdata; access requires a signed confidentiality and security agreement with the SRI. Contact the corresponding author (Section 2) or the SRI directly. | April 2022                                                        | Servicio de Rentas Internas (SRI), Ecuador. Administrative tax records (confidential, restricted access).                                                                                                            | Restricted to SRI facilities; academic research use only under Ecuador’s national transparency law; de-identified extracts must be deleted after replication (see NOTE in Section 2).       |
| cpi_year.dta                                                                                                                 | Instituto Nacional de Estadística y Censos (INEC), Ecuador – Índice de Precios al Consumidor (IPC)          | <a href="https://www.ecuadorencifras.gob.ec/indice-de-precios-al-consumidor-ipc/">https://www.ecuadorencifras.gob.ec/indice-de-precios-al-consumidor-ipc/</a>                                                  | April 2022                                                        | Instituto Nacional de Estadística y Censos (INEC). <i>Índice de Precios al Consumidor (IPC)</i> . Quito: INEC. Compiled into Stata format by Alex Bajaña.                                                            | Public data (INEC); <b>cpi_year.dta</b> compiled by Alex Bajaña.                                                                                                                            |
| icij_all.dta                                                                                                                 | ICIJ – Offshore Leaks Database (original leak data); SRI, Ecuador (taxpayer registry match & anonymization) | Original ICIJ data: <a href="https://offshoreleaks.icij.org/pages/database">https://offshoreleaks.icij.org/pages/database</a> . <b>icij_all.dta</b> itself is not publicly available (see notes).              | 5 March 2022 (ICIJ download); RUC registry downloaded 27 May 2022 | ICIJ. <i>Offshore Leaks Database, Panama Papers</i> . Washington, DC: ICIJ (2016 release, accessed 5 March 2022). Matched to SRI taxpayer records and anonymized by SRI personnel.                                   | Original ICIJ data: Open Database License / CC BY-SA; ICIJ must always be cited. <b>icij_all.dta</b> links RUC identifiers to ICIJ records; it is confidential and cannot be redistributed. |
| países_mid_aps.dta,<br>países_clasificados.csv                                                                               | Servicio de Rentas Internas (SRI), Ecuador – official annual list of tax havens / preferential tax regimes  | Not linked directly in the code. Official list published via Resolución NAC-DGERCGC15-00000052 and its annual amendments (SRI regulatory resolutions).                                                         | April 2022 (list current as of the October 2022 update)           | Servicio de Rentas Internas (SRI), Ecuador. <i>Resolución NAC-DGERCGC15-00000052 y reformas anuales – Listado de paraísos fiscales y regímenes fiscales preferentes</i> . Compiled into Stata format by Alex Bajaña. | Public regulatory document (SRI); <b>países_mid_aps.dta</b> / <b>países_clasificados.csv</b> compiled by Alex Bajaña.                                                                       |

Table 3: Main intermediate datasets produced by 1\_replication\_build.do (confidential working directory)

| Dataset                                                                                                                                                                                                                                                                                                   | Description                                                                                                    | Notes        | Pro-<br>vided |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|--------------|---------------|
| aps_processed.dta                                                                                                                                                                                                                                                                                         | Cleaned, deduplicated APS panel with aps / aps_2014 / aps_2012_2014 filing indicators.                         | Confidential | No            |
| crosswalk_ruc_id_update.dta                                                                                                                                                                                                                                                                               | Filtered RUC crosswalk (excluded firms only).                                                                  | Confidential | No            |
| f101_active_list.dta,<br>f101_id_list.dta,<br>f101_processed.dta, f101_long.dta                                                                                                                                                                                                                           | F101 activity/revenue panel and firm-ID lists.                                                                 | Confidential | No            |
| f102_id_list.dta                                                                                                                                                                                                                                                                                          | Firm-ID list from F102.                                                                                        | Confidential | No            |
| core_sample_list.dta,<br>core_sample_aps.dta,<br>core_sample_aps_processed.dta,<br>main_sample_aps_processed.dta                                                                                                                                                                                          | Core/main estimation sample construction (APS $\times$ F101 activity $\times$ industry exclusion).             | Confidential | No            |
| mid_2012.dta, mid_2013.dta,<br>mid_2014.dta, mid_2015.dta,<br>mid_2016.dta, mid_2017.dta,<br>mid_2018.dta, mid_2019.dta,<br>mid_processed.dta                                                                                                                                                             | Collapsed MID transaction totals by firm/year/country/reason.                                                  | Confidential | No            |
| aps_f101_id_list.dta,<br>panel_aps_f101_filing.dta                                                                                                                                                                                                                                                        | Combined APS/F101 firm-ID list and filing panel.                                                               | Confidential | No            |
| main_panel.dta, core_panel.dta                                                                                                                                                                                                                                                                            | Final main/core firm-year panels (APS $\times$ F101 merge); direct inputs to most of the 12 analysis do-files. | Confidential | No            |
| geo_alt_1_sample_aps.dta,<br>geo_alt_2_sample_aps.dta,<br>geo_alt_3_sample_aps.dta,<br>geo_alt_1_sample_aps_processed.dta,<br>geo_alt_2_sample_aps_processed.dta,<br>geo_alt_3_sample_aps_processed.dta,<br>geo_alt_1_panel.dta,<br>geo_alt_2_panel.dta,<br>geo_alt_3_panel.dta,<br>countries_geo_alt.dta | Alternative geographic-classification panels (robustness).                                                     | Confidential | No            |

Table 4: Analysis-ready datasets produced by the 12 do-files in `replication_do_files/` (included in this package)

| Dataset                                                                                                                                                                                                                                                                                                                                   | Produced by                       | Notes                        | Pro-<br>vided |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|------------------------------|---------------|
| core_time_series.dta<br>core_time_series_all.dta<br>core_time_series_meta.dta<br>core_time_series_meta_all.dta<br>balance_core_v_excluded_noaps_collapsed.dta<br>extensivemargin_counts_aps_forwarded.dta<br>extensivemargin_counts_aps_forwarded_f101.dta<br>histograms_participation_prominent_bygroup_by_year.dta                      | descriptive_statistics_balance.do | 8 files, see §7              | Yes           |
| processed/tax_benchmarking_coresample.dta                                                                                                                                                                                                                                                                                                 | tax_benchmarking.do               | –                            | Yes           |
| first_stage.dta                                                                                                                                                                                                                                                                                                                           | first_stage_results.do            | –                            | Yes           |
| processed/f101_aggregates.dta<br>processed/geo_counts_raw_panel_a_core_aps.dta<br>processed/geo_counts_panel_a_core_aps.dta<br>processed/geo_counts_panel_b_core_aps.dta<br>processed/geo_counts_panel_c_core_aps.dta<br>processed/geo_counts_panel_a_core_aps_50alt.dta                                                                  | geo_counts.do                     | panels a/b/c + 50alt variant | Yes           |
| main_ownership_results.dta<br>main_ownership_results_weighted.dta <sup>1</sup>                                                                                                                                                                                                                                                            | main_ownership_results.do         | –                            | Yes           |
| processed/geo_alt_1_time_series.dta<br>processed/geo_alt_2_time_series.dta<br>processed/geo_alt_3_time_series.dta<br>geo_alt_1_results.dta<br>geo_alt_2_results.dta<br>geo_alt_3_results.dta<br>geo_alt_1_results_weighted.dta <sup>1</sup><br>geo_alt_2_results_weighted.dta <sup>1</sup><br>geo_alt_3_results_weighted.dta <sup>1</sup> | geo_alt_ownership_results.do      | –                            | Yes           |
| main_ownership_mean_reversion.dta                                                                                                                                                                                                                                                                                                         | mean_reversion.do                 | –                            | Yes           |
| balance_consistent.dta<br>main_results_consistent_balance.dta                                                                                                                                                                                                                                                                             | ex_post_verification_tests.do     | –                            | Yes           |
| ex_ante_heterogeneity.dta                                                                                                                                                                                                                                                                                                                 | ex_ante_heterogeneity.do          | –                            | Yes           |
| heterogeneity_icij_majors.dta                                                                                                                                                                                                                                                                                                             | icij.do                           | –                            | Yes           |
| continuous_exposure.dta                                                                                                                                                                                                                                                                                                                   | continuous_exposure.do            | –                            | Yes           |
| expost_results.dta                                                                                                                                                                                                                                                                                                                        | expost_results.do                 | –                            | Yes           |

Some datasets in Table 4 are written a second time, under the same filename, in two different locations. `3_replication_exhibits_construction.do` reads the version in `replication_exhibits_sourcefiles/`, applies the additional processing required to build the exhibits, and saves the result under `replication_exhibits_sourcefiles/processed/`. In these cases both copies are legitimate – one per stage of the pipeline – and are expected to appear as duplicate filenames across two folders in the

<sup>1</sup>Physically present in `replication_exhibits_sourcefiles/` in this package, but not yet matched against an entry in the consolidated data-hash report at the time of writing; pending re-hash/verification by the author team.

data-hash report.

In addition, `3_replication_exhibits_construction.do` produces a further set of intermediate “\_tables.dta”/“datasets under `replication_exhibits_sourcefiles/processed/`. These are one processing step upstream of the final `.tex/.png` exhibits, and are also included in this package:

Table 5: Second-stage intermediate datasets produced by `3_replication_exhibits_construction.do` (included in this package)

| <b>Dataset</b><br><i>(in replication_exhibits_sourcefiles/processed/)</i> | <b>Built from</b>                           |
|---------------------------------------------------------------------------|---------------------------------------------|
| balance_coresample.dta                                                    | processed/tax_benchmarking_coresample.dta   |
| balance_core_v_excluded_noaps.dta                                         | balance_core_v_excluded_noaps_collapsed.dta |
| main_ownership_results_tables.dta                                         | main_ownership_results.dta                  |
| geo_alt_1_tables.dta                                                      | geo_alt_1_results.dta,                      |
| geo_alt_2_tables.dta                                                      | geo_alt_2_results.dta,                      |
| geo_alt_3_tables.dta                                                      | geo_alt_3_results.dta                       |
| heterogeneity_icij_majors_ddd_tables.dta                                  | heterogeneity_icij_majors.dta               |
| continuous_exposure_all_tables.dta                                        | continuous_exposure.dta                     |
| expost_results_tables.dta                                                 | expost_results.dta                          |
| main_ownership_mean_reversion_tables.dta                                  | main_ownership_mean_reversion.dta           |
| balance_consistent_tables.dta                                             | balance_consistent.dta                      |
| main_results_consistent_balance_table.dta                                 | main_results_consistent_balance.dta         |
| ex_ante_heterogeneity_graphs.dta                                          | ex_ante_heterogeneity.dta                   |

### 3.2 Orbis Data

This section describes every dataset used for replicating the results using Orbis data. Table 6 displays the raw Orbis source files. Table 7 provides details on raw datasets. Table 8 lists the analysis-ready datasets. All data from Orbis is not publicly available, and thus not included in the public version of the reproducibility package. The authors had access to the data through a user in Orbis website. This user requires paying a subscription fee.

Table 6: Raw datasets

| Dataset                            | Description                                                                                                                                         | Notes                                                                                                     | 1st<br>Access | Ac-<br>cess | Pro-<br>vided |
|------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|---------------|-------------|---------------|
| Export_GUO_New_Alt_<year>.xlsx     | Datasets with global ultimate owner with <year> spanning between 2016 and 2024.                                                                     | Confidential                                                                                              | Sept 2025     | No          |               |
| Export_EC_bvd_*.xlsx               | Datasets with information of subsidiaries and direct ownership.                                                                                     | Confidential                                                                                              | Sept 2025     | No          |               |
| flat-ui___data-Fri Sep 12 2025.csv | Data file with a list of ISO code 2 digit at the country level.                                                                                     | Public<br><a href="https://github.com/datasets/country-list">https://github.com/datasets/country-list</a> | Sept 2025     | No          |               |
| havens_2014.xls                    | List at the country level identifying tax havens as of SRI definition in 2014.                                                                      | Public through SRI resolution NAC-DGERCGC15-00000052                                                      | Sept 2025     | No          |               |
| NACE2_CBF_Table.xlsx               | List with NACE Rev. 2 (Statistical Classification of Economic Activities in the European Community), published by Eurostat on the EU's data portal. | Public<br><a href="http://data.europa.eu/ux2/nace2">http://data.europa.eu/ux2/nace2</a>                   | Sept 2025     | No          |               |

Table 7: Data source details for the raw datasets in Table 6

| Filename(s)                       | Source                                                                                                  | URL / access instructions                                                                                                                                                                    | Access date    | Citation                                                                                                                                                                                   | License                                                                                                                                                                                                                                               |
|-----------------------------------|---------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Export_GUO_New_Alt_<year>.xlsx    | Orbis by Bureau van Dijk/Moodys                                                                         | Not publicly available. Data accessible only through <a href="#">subscription</a> .                                                                                                          | September 2025 | Orbis (2025). Datasets with global ultimate owner between 2016 and 2024.                                                                                                                   | The authors accessed the Orbis data through a World Bank user account under a paid subscription. This subscription and user account are no longer active.                                                                                             |
| Export_EC_bvd_*.xlsx              | Orbis by Bureau van Dijk/Moodys                                                                         | Not publicly available. Data accessible only through <a href="#">subscription</a> .                                                                                                          | September 2025 | Orbis (2025). Datasets with subsidiaries and direct ownership.                                                                                                                             | The authors accessed the Orbis data through a World Bank user account under a paid subscription. This subscription and user account are no longer active.                                                                                             |
| flat-ui__data-Fri Sep 12 2025.csv |                                                                                                         | Publicly available through <a href="https://github.com/datasets/country-list">https://github.com/datasets/country-list</a> . Data file with a list of ISO code 2 digit at the country level. | September 2025 | Open Knowledge Foundation. <i>Country List</i> . GitHub repository, <a href="#">datasets/country-list</a> .                                                                                | Public Domain Dedication and License (PDDL), as applied by the maintainers of the GitHub repository. Note: the underlying data is ultimately sourced from ISO 3166-1, whose own licensing terms restrict use to internal and non-commercial purposes. |
| havens_2014.xls                   | Servicio de Rentas Internas (SRI), Ecuador official annual list of tax havens/preferential tax regimes. | Not linked directly in the code. Official list published via Resolución NAC-DGERCGC15-00000052 and its annual amendments (SRI regulatory resolutions).                                       | April 2022     | Servicio de Rentas Internas (SRI), Ecuador. Resolución NAC-DGERCGC15-00000052 y reformas anuales Listado de paraísos fiscales y regímenes fiscales preferentes.                            | Public regulatory document.                                                                                                                                                                                                                           |
| NACE2_CBF_Table.xlsx              | Eurostat                                                                                                | <a href="http://data.europa.eu/ux2/nace2">http://data.europa.eu/ux2/nace2</a>                                                                                                                | Sept 2025      | Eurostat. <i>NACE Rev. 2 Statistical Classification of Economic Activities in the European Community</i> . Luxembourg: Office for Official Publications of the European Communities, 2008. | Public access through Eurostat, EU's data portal.                                                                                                                                                                                                     |

Table 8: Processed datasets

| Dataset                       | Description                                                                                                                                                                                            | Notes        | 1st Access | Provided |
|-------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|------------|----------|
| Export_EC_bvd.xlsx            | Cleaned data file including subsidiaries and direct ownership.                                                                                                                                         | Confidential | Sept 2025  | No       |
| panel_tax_haven.dta           | Balanced skeleton panel at the subsidiary $\times$ year $\times$ owner-country-type level (non-tax haven, tax haven, unknown/Ecuador), for 2014–2019. Used to impute missing post-reform ownership.    | Confidential | Sept 2025  | No       |
| panel.dta                     | Balanced skeleton panel at the subsidiary $\times$ year level for 2014–2019, restricted to firms present in 2014.                                                                                      | Confidential | Sept 2025  | No       |
| exposure_definition.dta       | Cross-section of 2014 ownership classifications defining treatment and control groups: mostly haven-owned ( $\geq 50\%$ ), any haven ownership, non-haven owned, and majority foreign non-haven owned. | Confidential | Sept 2025  | No       |
| working_data.dta              | Main analysis panel (2014–2019) of Ecuadorian subsidiaries active in 2014, with treatment indicators, baseline ownership outcomes, in-sample flags, and the post-reform dummy.                         | Confidential | Sept 2025  | No       |
| balanced_sample_2014_2018.dta | Balanced two-period (2014 vs. 2018) sample of firms observed in both years, collapsed to the firm level with indicators for whether direct ownership changed.                                          | Confidential | Sept 2025  | No       |
| imputed_sample_2014_2018.dta  | Two-period (2014 vs. 2018) sample with post-reform missing ownership values carried forward from the last available year, including attrition indicators.                                              | Confidential | Sept 2025  | No       |

## 4 Computational Requirements

### 4.1 Software requirements

Required software is **Stata**, version 18. The license is shared among SRI researchers and was provided by the SRI for this project.

### 4.2 Memory and runtime requirements

#### 4.2.1 SRI administrative data

The code was last run on a machine with an **AMD Ryzen 7 1800X (8-core) processor and 64 GB of RAM**, running **Windows 10**. The active configuration block in each master do-file (labelled “SRI ENVIRONMENT”, e.g. line 22 of `1_replication_build.do`) points to local paths on the SRI’s own computing environment (`D:\BM_EXTENSION\B202106_JAKOB_BROUNSTEIN`); this is consistent with the run having been executed on that environment rather than a personal laptop.

Based on the timer output in the most recent logs (run of 10 July 2026), the full pipeline took approximately:

- `1_replication_build.do`: 5,016 seconds ( $\approx$  83.6 minutes)
- `2_replication_analysis.do`:  $\approx$  61.0 minutes (11:10:06 to 12:11:13)
- `3_replication_exhibits_construction.do`: 30.6 seconds

For a total of approximately **145 minutes ( $\approx$  2.4 hours)** to run the full `replication_masterdo.do`. All three files completed without any Stata errors in this run.

**Note on packages.** The code installs `ftools` and `gtools` via `net install` from local institutional paths (e.g. `F:\DTO_ESTUDIOS_E1_ext2\...`), not via `ssc install`, suggesting the SRI environment does not have direct internet/SSC access. Replicators running the code outside that environment should replace these lines with the standard `ssc install ftools / ssc install gtools` commands.

#### 4.2.2 Orbis data

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 2 minutes.

## 5 Instructions to replicators

### 5.1 Folder structure: SRI administrative data

This replication package uses the following folder structure, as set up by the `glob` macros in the three master do-files:

- `do/ecuador_transparency_replication/` – code:
  - `1_replication_build.do`, `2_replication_analysis.do`,
  - `3_replication_exhibits_construction.do`, `replication_masterdo.do`,
  - and the 12 do-files in `replication_do_files/`
- `data/transparency/replication/` – `$datadir`: confidential intermediate datasets built by `1_replication_build.do`

- `replication/logs/` – `$logdir`: one `.smcl` log per master do-file, per run date
- `replication/replication_exhibits_sourcefiles/` – `$outdir / $basedir`: analysis-ready datasets, included in this package
- `replication/replication_main_exhibits/graphs/` and `/tables/` – `$graphdir / $tabledir`: final `.png` figures and `.tex` tables, included in this package

The codes described in Section 7 rely on this folder structure via the global macros defined at the top of each master do-file.

## 5.2 Before replication: SRI administrative data

Each of the three master do-files contains several commented-out alternative `glob localdir` blocks (one per team member’s local machine, labelled e.g. “JAKOB ENVIRONMENT”, plus the active “SRI ENVIRONMENT” block) – only one block should be uncommented at a time. To replicate outside of the SRI environment, the replicator must:

- Edit `glob localdir` to point to the local project root. It is currently set at:  
line 22 of `1_replication_build.do`  
line 33 of `2_replication_analysis.do`  
line 56 of `3_replication_exhibits_construction.do`
- Edit `glob dodir` to point to the location of the do-files (line 45 of `2_replication_analysis.do`).
- Edit line 1 of `replication_masterdo.do` (the `cd` path) if running the three files through the master do-file.
- Remove or adjust the `sysdir set PLUS / PERSONAL` lines (institution-specific ado-file paths) and replace the `net install ftools/gtools ... from(...)` lines with `ssc install ftools / ssc install gtools`, since the current lines point to internal SRI network drives.

If the replicator obtains access to the de-identified SRI data, the raw and intermediate datasets listed in Table 1 and Table 3 would need to be placed at the paths referenced above for `1_replication_build.do` to run from scratch. Replication starting from the analysis-ready datasets in Table 4 (already included in this package) only requires `2_replication_analysis.do`’s outputs to be in place and can skip straight to `3_replication_exhibits_construction.do`.

## 5.3 Folder structure: Orbis data

For full replication of the results using Orbis data, we suggest the following, complete folder structure.

```
1_data
    1_Raw data
    2_Aux data
    3_Processed
2_codes
3_output
    2_figures
    3_tables
```

We provide the do-files to create the datasets used in the project and the subsequent analysis (from `1_Raw` data and after processing are stored in `1_Processed`). 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.

#### 5.4 Before replication: Orbis data

Before running the code and to make the replication process more efficient, in line 50 in the do-file `Ob_master.do`, the reviewer must insert the user name and the location where raw data and output files are located. Cleaning of raw data is run in R, thus requiring the command `rscript` is required in order to be able to run this code in Stata.

## 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 9 provides a mapping between all tables of the paper (including the online appendix) and the codes producing these results. All 22 tables using SRI data (from table 1 to table A15) are generated by the single file `3_replication_exhibits_construction.do`, via the `dataout` command. Filenames below match `replication_main_exhibits/tables/` exactly. “...” in the Code column refers to `3_replication_exhibits_construction.do`. Tables B1 and B2 are generated using Orbis data, and can be found in the folder `Orbis_replication/Code`

Table 9: Mapping of tables and analysis code

| Exhibit name | Output filename                                                                                                            | Code (line)                                      |
|--------------|----------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|
| Table 1      | <code>table_1_dd_shareholdership.tex</code>                                                                                | ...:671                                          |
| Table 2      | <code>table_2_dd_shareholdership_haven_dividends.tex</code>                                                                | ...:799                                          |
| Table 3      | <code>table_3_dd_continuous_exposure.tex</code>                                                                            | ...:930                                          |
| Table 4      | <code>table_4_dd_financials_profit_tax.tex</code>                                                                          | ...:1052                                         |
| Table 5      | <code>table_5_dd_financials_inputs.tex</code>                                                                              | ...:1177                                         |
| Table 6      | <code>table_6_dd_expost.tex</code>                                                                                         | ...:1338                                         |
| Table 7      | <code>table_7_tax_benchmarking_coresample_short.tex</code>                                                                 | ...:1444                                         |
| Table A1     | <code>table_a1_covariate_balance_excluded_firms_aps.tex</code>                                                             | ...:2053                                         |
| Table A2     | <code>table_a2_panel_balance_all.tex</code>                                                                                | ...:2071                                         |
| Table A3     | <code>table_a3_core_sample_descriptive_statistics.tex</code>                                                               | ...:2164                                         |
| Table A4     | <code>table_a4_taxhavens_geography_2014_merge_core_aps.tex</code>                                                          | ...:2474                                         |
| Table A5     | <code>table_a5_balance_table_major_domestic.tex</code>                                                                     | ...:2650                                         |
| Table A6     | <code>table_a6_dd_shareholdership_panel_b.tex</code>                                                                       | ...:2770                                         |
| Table A7     | <code>table_a7_dd_shareholdership_additional_mechanisms.tex</code>                                                         | ...:2894                                         |
| Table A8     | <code>table_a8_dd_post_tax_profit_allocation.tex</code>                                                                    | ...:3021                                         |
| Table A9     | <code>table_a9_dd_compliance_panelbalance.tex</code>                                                                       | ...:3139                                         |
| Table A10    | <code>table_a10_panel_a_robustness_geo_alt.tex,</code><br><code>table_a10_panel_b_main_ownership_mean_reversion.tex</code> | ...:3410, 3527                                   |
| Table A11    | <code>table_a11_ex_ante_heterogeneity.tex</code>                                                                           | ...:3714                                         |
| Table A12    | <code>table_a12_ddd_heterogeneity_ex_post_verification_test.tex</code>                                                     | ...:3853                                         |
| Table A13    | <code>table_a13_icij.tex</code>                                                                                            | ...:3989                                         |
| Table A14    | <code>table_a14_dd_continuous_alt_outcomes.tex</code>                                                                      | ...:4185                                         |
| Table A15    | <code>table_a15_dd_productive_inputs_weighted.tex</code>                                                                   | ...:4307                                         |
| Table B1     | <code>sample_2014_subsidiaries.tex</code>                                                                                  | <code>2a_descriptives_direct_ownership.do</code> |
| Table B2     | <code>diff_in_diff_main_tax_haven_owned.tex</code>                                                                         | <code>3a_analysis_diff_in_diff.do</code>         |

## 6.2 Mapping of figures and analysis code

Table 10 provides a mapping between figures and the code producing them. All figures are generated by `3_replication_exhibits_construction.do` via `graph export`. Filenames are the actual filenames found in `replication_main_exhibits/graphs/`. Figures A1, A2, and C1 (referenced in the paper’s appendix) are **not** produced by any code in this replication package and are not included in `replication_main_exhibits/graphs/`; they have been removed from this mapping pending confirmation from the author team of where they are generated. For Figure A5, both panels are exported by the code with an identical `_control` suffix (line 1597 for panel a, line 1671 for panel b) – filenames are reproduced here exactly as they appear in the code, without assuming a treatment/control distinction that the code itself does not encode.

Table 10: Mapping of figures and analysis code

| Exhibit Name | Panels | Output filename                                                                                                                                                                                                                                                                                                    | Code (line)            |
|--------------|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| Figure 1     | a-d    | figure_1a_firms_dd_shareholdership_timeseries.png<br>figure_1b_firms_dd_participation_prominent_levels_majors_twfe_unweighted.png<br>figure_1c_timeseries_bf_persona_timeseries.png<br>figure_1d_firms_dd_bf_persona_levels_majors_twfe_unweighted.png                                                             | 388, 402, 414, 427     |
| Figure 2     | a,b    | figure_2a_firms_dd_ownershipshares_alt_all_majors_twfe_unweighted.png<br>figure_2b_firms_dd_any_presence_in_havens.png                                                                                                                                                                                             | 460, 478               |
| Figure 3     | a-d    | figure_3a_firms_dd_cit_liability_w_log_majors_twfe_unweighted.png<br>figure_3b_firms_dd_gross_profit_w_log_majors_twfe_unweighted.png<br>figure_3c_firms_dd_cit_liability_w_binary_majors_twfe_unweighted.png<br>figure_3d_firms_dd_gross_return_on_assets_alt_w_majors_twfe_unweighted.png                        | 497, 512, 527, 543     |
| Figure A3    |        | figure_a3_histogram_control_exposure_2014.png                                                                                                                                                                                                                                                                      | 1478                   |
| Figure A4    | a-d    | figure_a4_panel_a_first_stage_tasa_ir_base.png<br>figure_a4_panel_b_first_stage_tasa_ir_base_aps_2015_to_2019.png<br>figure_a4_panel_c_first_stage_tasa_ir_base_cc_aps_2015_to_2019_no_extras.png<br>figure_a4_panel_d_first_stage_tasa_ir_base_cc_f101_2015_to_2019_no_extras.png                                 | 1495, 1508, 1521, 1533 |
| Figure A5    | a,b    | figure_a5_panel_a_balance_consistent_2015_2017_2014_control.png<br>figure_a5_panel_b_balance_consistent_2015_2017_2014_control.png                                                                                                                                                                                 | 1597, 1671             |
| Figure A6    | a,b    | figure_a6_panel_a_firms_dd_participation_prominent_log_majors_twfe_unweighted.png<br>figure_a6_panel_b_firms_dd_participation_prominent_poisson_majors_twfe_unweighted.png                                                                                                                                         | 1685, 1695             |
| Figure A7    |        | figure_a7_firms_dd_mid_div_salidas_prom_w_binary_majors_twfe_unweighted.png                                                                                                                                                                                                                                        | 1708                   |
| Figure A8    | a,b    | figure_a8_panel_a_extensive_margin_time_series_counts_operating_majors.png<br>figure_a8_panel_b_extensive_margin_time_series_assets_operating_majors.png                                                                                                                                                           | 1730, 1748             |
| Figure A9    | a,b    | figure_a9_panel_a_firms_dd_porcentaje_pff_levels_majors_twfe_unweighted.png<br>figure_a9_panel_b_firms_dd_porcentaje_ext_levels_majors_twfe_unweighted.png                                                                                                                                                         | 1765, 1778             |
| Figure A10   | a-c    | figure_a10_panel_a_firms_dd_b_mid_salidas_pff_2014_participation_prominent_levels_majors_twfe_unweighted.png<br>figure_a10_panel_b_firms_dd_participation_bo_2014_treatment_group_med_majors_twfe_unweighted.png<br>figure_a10_panel_c_firms_dd_participation_prominent_levels_ex_ante_profits_twfe_unweighted.png | 1778, 1824, 1843       |
| Figure A11   | a-c    | figure_a11_panel_a_firms_dd_total_assets_alt_log_majors_twfe_unweighted.png<br>figure_a11_panel_b_firms_dd_investments_w_binary_majors_twfe_unweighted.png<br>figure_a11_panel_c_firms_dd_labor_cost_w_log_majors_twfe_unweighted.png                                                                              | 1862, 1877, 1892       |

## 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. Section 7.3 describes the codes producing the results with Orbis data. Some of the code relies on programs downloaded from the Statistical Software Components (SSC) archive. These programs are described in section 7.4.

Global paths to the different folders of the project (input files, output tables, figures, etc.) are established in the respective master do-files. 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 Cleaning data process

All data-creation steps are contained in a single file, [1\\_replication\\_build.do](#) (1,429 lines), organized into the ten blocks below (section headers as they appear in the code, with their approximate line ranges).

- [1\\_replication\\_build.do – PRE-PROCESSING](#) (lines 57–89)

Copies the CPI, ICIJ, country-classification, and RUC-crosswalk raw files into the working data directory.

Inputs: `cpi_year.dta`

`icij_all.dta`

`países_mid_aps.dta`

`países_clasificados.csv`

Outputs: `$datadir/cpi_year.dta`

`icij_all.dta`

`países_mid_aps.dta`

`países_clasificados.csv`

---

- [1\\_replication\\_build.do – APS UNLOADING + PROCESSING](#) (lines 90–145)

Reads and appends the yearly APS csv files, deduplicates, and builds firm-level filing indicators.

Inputs: `A_APS_<2012-2019>.csv`

Outputs: `$datadir/aps_processed.dta`

---

- [1\\_replication\\_build.do – LIST OF EXCLUDED FIRMS](#) (lines 146–158)

Filters the RUC crosswalk to identify government entities and oil companies excluded from the transparency law.

Inputs: `crosswalk_ruc_id_update_16082021.dta`

Outputs: `$datadir/crosswalk_ruc_id_update.dta`

---

- [1\\_replication\\_build.do – F101 UNLOADING + PROCESSING](#) (lines 159–214)

Reads and appends the yearly F101 csv files and builds revenue-based activity flags.

Inputs: F101\_<2012-2019>.csv  
Outputs: \$datadir/f101\_active\_list.dta  
f101\_id\_list.dta

---

- [1\\_replication\\_build.do – F102 RUC LIST \(lines 215–224\)](#)

Extracts a unique firm-ID list from the F102 source.

Inputs: f102\_processed\_all.dta  
Outputs: \$datadir/f102\_id\_list.dta

---

- [1\\_replication\\_build.do – BROAD APS X F101 ACTIVITY X INDUSTRY EXCLUSION & SAMPLE SET / ATTACHING THE CORE SAMPLE TO THE APS AND PROCESSING \(lines 225–781\)](#)

Builds the core and main estimation samples by combining APS filing status, F101 activity, and the excluded-firm list; processes the full F101 panel.

Inputs: \$datadir/aps\_processed.dta  
f101\_active\_list.dta  
f101\_id\_list.dta  
f102\_id\_list.dta  
crosswalk\_ruc\_id\_update.dta  
Outputs: \$datadir/core\_sample\_list.dta  
core\_sample\_aps.dta  
core\_sample\_aps\_processed.dta  
main\_sample\_aps\_processed.dta  
f101\_processed.dta  
f101\_long.dta  
panel\_aps\_f101\_filing.dta  
aps\_f101\_id\_list.dta

---

- [1\\_replication\\_build.do – MID \(lines 782–888\)](#)

Reads and collapses the yearly MID (international currency-transfer) csv files by firm, transaction type, country, and reason code, restricted to firms in the APS/F101 sample.

Inputs: A\_MID\_<2012-2019>.csv  
\$datadir/aps\_f101\_id\_list.dta  
Outputs: \$datadir/mid\_2012.dta  
mid\_2013.dta  
mid\_2014.dta  
mid\_2015.dta  
mid\_2016.dta  
mid\_2017.dta  
mid\_2018.dta  
mid\_2019.dta  
mid\_processed.dta

- 
- [1\\_replication\\_build.do](#) – COMBINE ALL OF THE FILES AND PROCESS (lines 889–1003)  
Merges the main/core APS samples with the processed F101 panel to build the two final firm-year panels used throughout the analysis code.

Inputs: \$datadir/main\_sample\_aps\_processed.dta  
core\_sample\_aps\_processed.dta  
f101\_processed.dta  
Outputs: \$datadir/main\_panel.dta  
core\_panel.dta

---

- [1\\_replication\\_build.do](#) – GEOGRAPHY ROBUSTNESS BUILD (lines 1004–1409)  
Repeats the sample-construction and merge steps under three alternative geographic classifications (haven vs. Colombia; all havens vs. Colombia; all havens vs. all Latin American non-havens), for robustness checks.

Inputs: \$datadir/core\_sample\_aps.dta  
paises\_mid\_aps.dta  
f101\_processed.dta  
Outputs: \$datadir/geo\_alt\_1\_sample\_aps.dta  
geo\_alt\_2\_sample\_aps.dta  
geo\_alt\_3\_sample\_aps.dta  
geo\_alt\_1\_sample\_aps\_processed.dta  
geo\_alt\_2\_sample\_aps\_processed.dta  
geo\_alt\_3\_sample\_aps\_processed.dta  
geo\_alt\_1\_panel.dta  
geo\_alt\_2\_panel.dta  
geo\_alt\_3\_panel.dta  
countries\_geo\_alt.dta

---

## 7.2 Empirical analysis

### 7.2.1 Analysis using SRI data

All analysis code is orchestrated by [2\\_replication\\_analysis.do](#), which calls the following 12 do-files (in `replication_do_files/`), in this order, each reading from the confidential panels in `$datadir` ([Table 3](#)) and writing to `replication_exhibits_sourcefiles/` ([Table 4](#), included in this package).

- [descriptive\\_statistics\\_balance.do](#)  
Descriptive statistics, time series, and covariate-balance checks for the core sample.  
Inputs: \$datadir/aps\_processed.dta  
\$datadir/core\_panel.dta  
\$datadir/f101\_processed.dta  
Outputs: processed/core\_time\_series.dta

processed/core\_time\_series\_all.dta  
processed/core\_time\_series\_meta.dta  
processed/core\_time\_series\_meta\_all.dta  
balance\_core\_v\_excluded\_noaps\_collapsed.dta  
extensivemargin\_counts\_aps\_forwarded.dta  
extensivemargin\_counts\_aps\_forwarded\_f101.dta  
histograms\_participation\_prominent\_bygroup\_by\_year.dta

---

- [tax\\_benchmarking.do](#)

Benchmarks reported tax variables against the core sample.

Inputs: \$datadir/aps\_processed.dta

\$datadir/core\_panel.dta

Outputs: processed/tax\_benchmarking\_coresample.dta

---

- [first\\_stage\\_results.do](#)

Estimates the first-stage regressions of the exposure/treatment measure.

Inputs: \$datadir/main\_panel.dta

Outputs: first\_stage.dta

---

- [geo\\_counts.do](#)

Builds geographic count aggregates for the core sample.

Inputs: \$datadir/core\_sample\_list.dta

Outputs: processed/f101\_aggregates.dta

processed/geo\_counts\_raw\_panel\_a\_core\_aps.dta

processed/geo\_counts\_panel\_a\_core\_aps.dta

processed/geo\_counts\_panel\_b\_core\_aps.dta

processed/geo\_counts\_panel\_c\_core\_aps.dta

processed/geo\_counts\_panel\_a\_core\_aps\_50alt.dta

---

- [main\\_ownership\\_results.do](#)

Main difference-in-differences ownership regressions (unweighted and weighted).

Inputs: \$datadir/main\_panel.dta

Outputs: main\_ownership\_results.dta

main\_ownership\_results\_weighted.dta (*pending hash-report verification, see [Table 4](#)*)

---

- [geo\\_alt\\_ownership\\_results.do](#)

Repeats the ownership regressions under the three alternative geographic classifications.

Inputs: \$datadir/geo\_alt\_1\_panel.dta

\$datadir/geo\_alt\_2\_panel.dta

\$datadir/geo\_alt\_3\_panel.dta

Outputs: processed/geo\_alt\_1\_time\_series.dta

processed/geo\_alt\_2\_time\_series.dta

processed/geo\_alt\_3\_time\_series.dta

geo\_alt\_1\_results.dta

geo\_alt\_2\_results.dta

geo\_alt\_3\_results.dta

geo\_alt\_1\_results\_weighted.dta

geo\_alt\_2\_results\_weighted.dta

geo\_alt\_3\_results\_weighted.dta (*all three pending hash-report verification, see Table 4*)

---

- [mean\\_reversion.do](#)

Tests for mean reversion in ownership shares.

Inputs: \$datadir/main\_panel.dta

Outputs: main\_ownership\_mean\_reversion.dta

---

- [ex\\_post\\_verification\\_tests.do](#)

Ex-post verification / consistent-balance robustness tests.

Inputs: \$datadir/core\_panel.dta

Outputs: balance\_consistent.dta

main\_results\_consistent\_balance.dta

---

- [ex\\_ante\\_heterogeneity.do](#)

Heterogeneity analysis by ex-ante (pre-reform) firm characteristics.

Inputs: \$datadir/main\_panel.dta

Outputs: ex\_ante\_heterogeneity.dta

---

- [icij.do](#)

Heterogeneity analysis by presence in the ICIJ offshore-leaks data.

Inputs: \$datadir/panel\_aps\_f101\_filing.dta

Outputs: heterogeneity\_icij\_majors.dta

---

- [continuous\\_exposure.do](#)

Regressions using a continuous (rather than binary) exposure/treatment measure.

Inputs: \$datadir/main\_panel.dta

Outputs: continuous\_exposure.dta

---

- [expost\\_results.do](#)

Ex-post firm outcome regressions (financials, profit allocation).

Inputs: `$datadir/main_panel.dta`

Outputs: `expost_results.dta`

---

### 7.2.2 Supplementary code: ICIJ list construction (run outside the SRI environment)

The following R scripts were used to construct `icij_all.dta` before it was anonymized and delivered by SRI personnel (see [section 2](#) for the full narrative). They were run on the authors' personal computers, not inside the SRI environment, in the ICIJ list creation project shared under `SRI_replication_11092026/`. They are included here for transparency on the construction of this input, not as part of the executable `ecuador_transparency_replication` pipeline.

- [Functions.R](#) (last modified 31 May 2022)  
Defines the deterministic (`f_precision_deter`) and fuzzy (`f_precision_fuzzy`, via `f_bid`) name-matching functions used by `01_Match_with_RUC.R`.
- 

- [01\\_Match\\_with\\_RUC.R](#) (last modified 28 June 2022)  
For each of the three Ecuador-restricted ICIJ files (entities, officers, intermediaries), performs a deterministic exact-name join against the SRI taxpayer registry, followed by a fuzzy join (cosine distance for officers/intermediaries; Jaro-Winkler distance, at increasing thresholds, for entities) on the names left unmatched.

Inputs: `ruc_completo_establecimientos.rds`

`ecuador_icij_officer_names.xlsx`

`ecuador_icij_intermediaries.xlsx`

`ecuador_icij_entities_names.xlsx`

Outputs: `Results/officer_match.rds`

`Results/intermediaries_match.rds`

`Results/entities_match.rds`

---

- [02\\_Final\\_format.R](#) (last modified 28 June 2022)  
Joins each of the three ICIJ files back to its match results and to the corresponding RUC record, producing the final matched files.

Inputs: `ruc_completo_establecimientos.rds`

`ecuador_icij_officer_names.xlsx`, `Results/officer_match.rds`

`ecuador_icij_intermediaries.xlsx`, `Results/intermediaries_match.rds`

`ecuador_icij_entities_names.xlsx`, `Results/entities_match.rds`

Outputs: `Results/final_officer.rds`

`Results/final_intermediaries.rds`

`Results/final_entities.rds`

`Results/final_entities.xlsx`

---

### 7.2.3 Exhibit construction

- [3\\_replication\\_exhibits\\_construction.do](#) (4,365 lines)  
Single file that reads every dataset in `replication_exhibits_sourcefiles/` ([Table 4](#)) and produces all final exhibits: 22 tables via `dataout` ([Table 9](#)) and 30 figures via `graph export`

(Table 10). It also writes second-stage intermediate datasets to `replication_exhibits_sourcefiles/processed/` (Table 5).

Inputs: all datasets in `replication_exhibits_sourcefiles/` (see Table 4)

Outputs: `replication_main_exhibits/tables/*.tex` (22 files)

`replication_main_exhibits/graphs/*.png` (30 files)

second-stage intermediate datasets in `replication_exhibits_sourcefiles/processed/` (see Table 5)

### 7.3 Analysis using Orbis data

- [1a\\_cleaning\\_subsidiaries.R](#)

Append cross section data on subsidiaries in case we have more than one file from Orbis.

Input:

`raw/Export_EC_bvd_<year>.xlsx`

`raw/havens_2014.xls`

`raw/NACE2_CBF_Table.xlsx`

Output:

`processed/Export_EC_bvd.xlsx`

---

- [1b\\_prepare\\_working\\_dataset.do](#)

Prepare working data on DIRECT ownership from Orbis with Ecuadorian subsidiaries.

Input:

`processed/Export_EC_bvd.xlsx`

Output:

`processed/panel_tax_haven.dta`

`processed/panel.dta`

`processed/exposure_definition.dta`

`processed/working_data.dta`

`processed/balanced_sample_2014_2018.dta`

`processed/imputed_sample_2014_2018.dta`

---

- [2a\\_descriptives\\_direct\\_ownership.do](#)

Analyze data on DIRECT ownership from Orbis with Ecuadorian subsidiaries

Input:

`processed/working_data.dta`

Output:

`tables/sample_2014_subsidiaries.tex`

---

- [3a\\_analysis\\_diff\\_in\\_diff.do](#)

Causal analysis on direct ownership from Orbis with Ecuadorian subsidiaries.

Estimates difference-in-differences regressions.

Input:

processed/imputed\_sample\_2014\_2018.dta

processed/balanced\_sample\_2014\_2018.dta

Output:

tables/diff\_in\_diff\_main\_tax\_haven\_owned.tex

---

## 7.4 Programs

This list was built strictly from command usage found in the do-files (occurrence counts across `*.do` and `replication_do_files/*.do`), not from the original template. `ftools` and `gtools` are installed via `net install ... from(<local SRI path>)` at the top of several do-files rather than via `ssc install` (see the note in Section 4); the remaining packages below are used in the code but have no visible install command in this package, so the replicator should install them via `ssc install` before running.

- **gtools** (used ~6 times directly, plus via `gegen` / `gcollapse` / `gduplicates`: 157/23/18 occurrences)  
Mauricio Caceres Bravo, 2018. “GTOOLS: Stata module to provide a fast implementation of common group commands,” Statistical Software Components S458514, Boston College Department of Economics.
- **reghdfe** (111 occurrences)  
Correia, Sergio. 2017. “Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator.” Working Paper.
- **ftools** (used ~6 times directly; required by `reghdfe`)  
Sergio Correia, and Julian Reif. (2023). `sergiocorreia/ftools`: `ftools` 2.49.1 (08aug2023) (2.49.1). Zenodo.
- **ppmlhdfc** (20 occurrences – used for the Poisson specification in Figure A6, panel b)  
Sergio Correia, Paulo Guimarães, and Thomas Zylkin, 2019. “PPMLHDFE: Stata module for Poisson pseudo-likelihood regression with multiple levels of fixed effects,” Statistical Software Components S458622, Boston College Department of Economics, revised 07 Sep 2023. Also published as: Correia, S., Guimarães, P., and Zylkin, T. (2020). “Fast Poisson estimation with high-dimensional fixed effects.” *The Stata Journal*, 20(1), 95–115.
- **regsave** (136 occurrences – used throughout the 12 analysis do-files to save regression coefficients to the datasets in Table 4)  
Julian Reif, 2008. “REGSAVE: Stata module to save regression results to a Stata-formatted dataset,” Statistical Software Components S456964, Boston College Department of Economics, revised 24 Jan 2025.
- **dataout** (23 occurrences – used exclusively in `3_replication_exhibits_construction.do` to export all 22 final tables to `.tex`)  
Roy Wada, 2009. “DATAOUT: Stata module to export a dataset or tab-delimited file into various formats,” Statistical Software Components S457022, Boston College Department of Economics, revised 28 Oct 2009.
- **carryforward** (3 occurrences)  
David Kantor, 2007. “CARRYFORWARD: Stata module to carry forward previous observations,” Statistical Software Components S456951, Boston College Department of Economics, revised 18 Sep 2016.
- **estout**  
Ben Jann, 2004. “ESTOUT: Stata module to make regression tables,” Statistical Software Components S439301, Boston College Department of Economics, revised 13 Apr 2026.

- **erepost**

Ben Jann, 2007. “EREPOST: Stata module to repost the estimation results,” Statistical Software Components S456850, Boston College Department of Economics, revised 30 Jun 2025.

- **frmtable**

Gallup, John Luke. 2012. “A programmer’s command to build formatted statistical tables.” *The Stata Journal*, 12(4), 655–673.

- **distinct**

Gary Longton and Nicholas J. Cox, 2002. “DISTINCT: Stata module to display distinct values of variables,” Statistical Software Components S424201, Boston College Department of Economics, revised 21 Mar 2012.

The following package is installed from GitHub through the master do-file for Orbis analysis:

- **rscript**

Julian Reif, 2020. **reifjulian/rscript**: Stata module to call an R script from within Stata. Available at <https://github.com/reifjulian/rscript>.