Two decades of top income shares in Honduras

Code Repository Documentation

Giselle Del Carmen[*]
Santiago Garriga[†]
Wilman Nuñez[‡]
Thiago Scot[§]

November 2023

---

[*]Enodo S.A., Honduras. Email: gdelcarmen@enodohn.com

[†]Centro de Estudios en Finanzas Públicas (CEFIP), IIE-FCE, Universidad Nacional de La Plata, Argentina. Email: santiago.garriga@econo.unlp.edu.ar

[‡]Servicio de Administración de Rentas (SAR), Honduras. Email: wponce@sar.gob.hn

[§]Development Impact Evaluation, World Bank. Email: tscot@worldbank.org

Contents

1 Overview

The codes in this repository replicate the tables and figures from "Two decades of top income shares in Honduras", by Del Carmen, Garriga, Nuñez, and Scot. The replication folder contains the codes to go from the raw administrative data to the results in the paper.

This documentation is structured as follows. Section 2 describes the data sources and their availability. Section 3 describes the datasets used in the analysis. Section 4 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 administrative data for this project is owned by the Honduran tax authority, the Servicio de Administración de Rentas. The data were made available to us exclusively for the purpose of this research project through collaboration agreements between the corresponding author and the government agency.

While at the time of the writing we cannot share the data for other projects, we can provide access for replication purposes of this study (conditional on 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 the agencies:

2.3 Details on each Data Source

The analysis is based on de-identified administrative raw data provided by the Honduran tax authority, the Servicio de Administración de Rentas (SAR). The main datasets include information on rental and

capital gains, gender and other personal information (e.g., marital status or location), shareholder information linked to the firms to which the individuals have any participation, corporate income tax (CIT), distributed dividends at the individual level. Data is available mainly from 2003 to 2020.

Part of the analysis also relies on Honduran macro data from the World Inequality Database (WID).

We also use survey data which includes information on income and its composition at the household level.

3 Datasets

Dataset	Description	Notes	Provided
WID_data_HN.csv	World Inequality Database data.	Public	No
Nueva Hogar'year'_clean.dta	Clean survey data.	Confidential	No
Nueva Hogar'year'_clean_v2.dta	Clean survey data.	Confidential	No
inflation.dta	Inflation data.	Public	No
Merged_'year'.dta	Combined dataset of tax records and survey data used to summarize the merging process between household survey microdata and administrative tax records.	Confidential	No
Merged_NoCorp_'year'.dta	Similar to the previous dataset, but without taking into account corporate profits.	Confidential	No
MacroAggregatesWDI.dta	Clean World Inequality data.	Public	No
InequalityMeasures.dta	Aggregated dataset at the top income groups level with information on the share of total income corresponding to such groups using different sources of data (tax records, survey data, with/withou corporate profits).	Confidential	No
Merged_macro_'year'.dta	Processed dataset with corporate profits and adjusted for macroeconomic aggregates.	Confidential	No
survey_population.dta	Population data.	Confidential	No
WID_PPP.dta	PPP exchange rate data.	Public	No
TaxData_Analysis_'year'.dta	Clean dataset at the individual level with information tax liability and income.	Confidential	No
HHSurvey_analysis_'year'.dta	Processed household survey dataset ready for analysis.	Confidential	No
merging_diagnostic.dta	Processed dataset used to illustrate the tax-to-survey observation ratios across income levels.	Confidential	No

Dataset	Description	Notes	Provided
231121_wid_aggregates	`wid` package data on total income and its composition as of November 21st, 2023.	Public	No
231121_aptinc	`wid` package data on pre-tax national income as of November 21st, 2023.	Public	No
231121_xlcusp	`wid` package data on exchange rates as of November 21st, 2023.	Public	No
231121_sptinc	`wid` package data on the shares of pre-tax income that accrues to a percentile of the population as of November 21st, 2023.	Public	No
pip	`pip` package data.	Public	No
TaxData_Gpercentiles_'year'.dta	Tax and income dataset at the g-percentile level.	Confidential	No
expansion_gpercentile.dta	Dataset at the g-percentile level with weights used to expand income when withholding can't be observed.	Confidential	No
components_adjustment_gpercentiles.dta	Dataset at the g-percentile level with weights that used to recover estimated income compositions for years when withholding can't be observed.	Confidential	No
CIT_clean.dta	Corporate income tax data.	Confidential	No
DMR_113_FirmLevel.csv	Dataset on distributed dividends	Confidential	No
UndistributedProfits_relation.dta	Dataset at the shareholder-corporation-year level with profits, turnover, and taxes information.	Confidential	No

4 Computational Requirements

Software requirements

The data creation and analysis code is written in `Stata`. Version 16 is used.

Controlled randomness

Memory and runtime requirements

We suggest the use of a processor with at least 8 cores and a RAM of 16 GB. With those settings, the replicator can expect full data creation to take about 15 minutes. Around 25 minutes can be expected to run the full data analysis.

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_ analysis*.

1_rawdata

2_analysis

      – 1_dta

      – 2_codes

      – 3_output

          – 1_log

          – 2_figures

          – 3_tables

This folder structure is reflected in the do-file `master_file.do`, which we provide in the code repository. We provide the do-files to create the datasets used in the project and the subsequent analysis (*2_ analysis*). The `master_file.do` do-file define *globals* in Stata, which correspond to the above folder tree. The codes described in section 7 rely on these folder paths. Each data creation and analysis code calls `master_file.do` in the beginning of the do-file.

6 List of tables, figures and programs

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 1 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 content of the tables.

Table 1: Mapping of tables and analysis code

Table	Code
Table 1	11_CreateTableCounts.do
Table 2	3a_Inequality_MergeTaxSurvey_loop.do
Table 3	7c_wid_TableComparisons
Table 4	Appendix_TaxRecords_Extra.do
Table A1	*Not generated by code*
Table A2	07a_Inequality_YearAnalysis_MacroDistributions.do
Table A3	7c_wid_TableComparisons
Table A4	7c_wid_TableComparisons
Table B1	Aux_CorporateMacroAggregates.do
Table C1	08bMacroAggregate_Descriptives.do
Table C2	5a_MacroAdjustments.do
Table E1	Appendix_TaxRecords_Extra.do
Table F1	2a_AgriculturalIncomeAdjust.do

6.2 Mapping of figures and analysis code

Table 2 provides a mapping between all figures of the paper (including the online appendix) and the code producing these results. The codes generate *.pdf* or *.png* files containing the graphs.

Table 2: Mapping of figures and analysis code

Figure	Panels	Code
Figure 1	A, B	5a_MacroAdjustments.do
Figure 2		07a_Inequality_YearAnalysis_MacroDistributions.do
Figure 3	A, B	07a_Inequality_YearAnalysis_MacroDistributions.do
Figure 4		6b_Inequality_MeasuringIneq_Graphs.do
Figure 5		6b_Inequality_MeasuringIneq_Graphs.do
Figure 6	A, B	Appendix_TaxRecords_Extra.do
Figure 7		6b_Inequality_MeasuringIneq_Graphs
Figure 8	A, B	16_NewTopStability.do
Figure 9		Appendix_TaxRecords_Extra.do
Figure 10		Appendix_TaxRecords_Extra.do
Figure A1	A, B, C	PIP_Code.do
Figure A2	A, B	07a_Inequality_YearAnalysis_MacroDistributions.do
Figure A3		08_TopStability.do
Figure A4		Aux_MergingDiagnostic.do
Figure C1		1f_MacroAggregate.do
Figure C2		08bMacroAggregate_Descriptives.do
Figure C3	A, B, C, D, E	Aux_WID_DescribeAggregates.do
Figure E1	A, B	Appendix_TaxRecords.do
Figure E2		Appendix_TaxRecords_Extra.do
Figure E3	A, B	Appendix_TaxRecords_Extra.do
Figure E4	A, B	Appendix_TaxRecords_Extra.do
Figure G1		2f_Aux_ExpansionGperc.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 rely on programs downloaded from the Statistical Software Components (SSC) archive. These programs are described in section 7.4.

Every code in our project folder loads a setup-script as the first line of code. This file is titled `master_name.do` and is part of the code repository. Global paths to the different folders of the project (input files, output tables, logs, etc.) are established through this call. 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, `master_name.do` installs all packages needed in the project that are downloaded from SSC.

7.1 Creation of data sets for analysis

1a_RentalCapGains.do

This do-file creates RentalIncome.dta and CapitalGains.dta datasets, which are used as intermediate datasets in the 1d_Inequality_Clean_loop and 1e_Inequality_Clean_2003_2010 do files.

Input data:

    $databases/Inputs/RawTaxData/MostUpdated_Apr23/Alquileres.csv

    $databases/Inputs/RawTaxData/MostUpdated_Apr23/Ganacias_Capital.csv

Output data:

    $databases/Output/CleanTaxData/RentalIncome.dta

    $databases/Output/CleanTaxData/CapitalGains.dta

1b_GenderData.do

This do-file creates Gender_age_dataset.dta dataset, which has gender information at the individual level and is used as intermediate dataset in the 1d_Inequality_Clean_loop and 1e_Inequality_Clean_2003_2010 do files.

Input data:

    $databases/Inputs/RawTaxData/Desigualdad_Jan23/Base_Edad_Genero.csv

Output data:

    $databases/Output/CleanTaxData/Gender_age_dataset.dta

1cc_CIT_Clean.do

This do-file creates CIT_clean.dta dataset, which has corporate income tax information and is used as intermediate dataset in the 1c_UndistributedProfits do file.

Input data:

    $databases/Inputs/RawTaxData/ISR_PJ_v3.dta

Output data:

    $databases/Output/Auxiliary/CIT_clean.dta

1c_UndistributedProfits.do

This do-file creates the UndistributedProfits_relation.dta and the UndistributedProfits_individual datasets, which are used as intermediate datasets in the 1e_Inequality_Clean_2003_2010, 1d_Inequality_Clean_loop, and the Aux_CorporateMacroAggregates do files.

Input data:

    $databases/Inputs/AuxData/clean_shareholder_list_IndividualOnly.dta

    $databases/Inputs/RawTaxData/DMR_113_v2.csv

    $databases/Output/Auxiliary/CIT_clean.dta

Output data:

    $databases/Output/Auxiliary/UndistributedProfits_relation.dta

    $databases/Output/Auxiliary/UndistributedProfits_individual.dta

---

1d_Inequality_Clean_loop.do

This do-file creates the TaxData_'year'.dta datasets for the 2011-2020 period, which have inequality measures related data and are used as intermediate datasets in the 2d_Inequality_CreateVars_loop do file.

Input data:

    $databases/Inputs/RawTaxData/MostUpdated_Apr23/'year'.csv

    $databases/Output/CleanTaxData/CapitalGains.dta

    $databases/Output/CleanTaxData/RentalIncome.dta

    $databases/Output/Auxiliary/UndistributedProfits_individual.dta

    $databases/Output/CleanTaxData/Gender_age_dataset

Output data:

    $databases/Output/CleanTaxData/TaxData_'year'.dta

1d_Inequality_Clean_2003_2010.do

This do-file creates the TaxData_'year'.dta datasets for the 2003-2010 period, which have inequality measures related data and are used as intermediate datasets in the 2d_Inequality_CreateVars_loop do file.

Input data:

    $databases/Inputs/RawTaxData/MostUpdated_Apr23/'year'.csv

    $databases/Output/CleanTaxData/CapitalGains.dta

    $databases/Output/CleanTaxData/RentalIncome.dta

    $databases/Output/Auxiliary/UndistributedProfits_individual.dta

    $databases/Output/CleanTaxData/Gender_age_dataset

Output data:

    $databases/Output/CleanTaxData/TaxData_'year'.dta

2b_Inequality_PrepareSurveyData_loop.do

This do-file creates the HHSurvey_analysis_'year'.dta datasets, which are later used in several data creation do files as input files (20_AggregatesMacro, 2c_Inequality_SurveyPopulation, 3a_Inequality_MergeTaxSurvey_loop, 6a_Inequality_MeasuringIneq, and 4a_NoCorp_Inequality_MergeTaxSurvey_loop).

Input data:

$databases/Inputs/CleanSurveyData/Nueva Hogar'year'_clean_v2.dta

$databases/Inputs/CleanSurveyData/Nueva Hogar'year'_clean.dta

Output data:

$databases/Output/AnalysisSurveyData/HHSurvey_analysis_'year'.dta

---

2c_Inequality_SurveyPopulation.do

This do-file creates the survey_population dataset, which has information on population levels across years.

Input data:

$databases/Output/AnalysisSurveyData/HHSurvey_analysis_'year'.dta

Output data:

$databases/Output/Auxiliary/survey_population.dta

2d_Inequality_CreateVars_loop.do

This do-file creates the TaxData_Analysis_'year'.dta datasets, which have information on income and tax liability.

Input data:

$databases/Output/CleanTaxData/TaxData_'year'.dta

Output data:

$databases/Output/AnalysisTaxData/TaxData_Analysis_'year'.dta

---

2e_Inequality_CreateGpercentiles_loop.do

This do-file creates the TaxData_Gpercentiles_NoCorp_'year'.dta and the TaxData_Gpercentiles_'year'.dta datasets using using microdata on taxpayers in each year and computing average outcomes and counts for 37 g-percentiles.

Input data:

$databases/Output/AnalysisTaxData/TaxData_Analysis_'year'.dta

$databases/Output/Auxiliary/survey_population.dta

Output data:

$databases/Output/AnalysisTaxData/TaxData_Gpercentiles_NoCorp_'year'.dta

$databases/Output/AnalysisTaxData/TaxData_Gpercentiles_'year'.dta

---

2g_Aux_Gpercent_ShAdjust.do

This do-file creates the components_adjustment_gpercentiles.dta dataset, which has weights used recover estimated income compositions. The resulting dataset is later used as input in the 3a_Inequality_MergeTaxSurvey_loop and the 4a_NoCorp_Inequality_MergeTaxSurvey_loop do files.

Input data:

$databases/Output/AnalysisTaxData/TaxData_Gpercentiles_'year'.dta

$databases/Inputs/AuxData/inflation.dta

Output data:

$databases/Output/Auxiliary/components_adjustment_gpercentiles.dta

---

4_NoCorp_2f_Aux_ExpansionGperc.do

This do-file creates the expansion_gpercentile_NoCorp.dta dataset, which has weights used to expand the amount of income observed in years with missing data. The dataset is used in the 4a_NoCorp_Inequality_MergeTaxSurvey_loop do file.

Input data:

$databases/Output/AnalysisTaxData/TaxData_Analysis_'year'.dta

$databases/Inputs/Auxiliary/survey_population.dta

Output data:

$databases/Output/Auxiliary/expansion_gpercentile_NoCorp.dta

4a_NoCorp_Inequality_MergeTaxSurvey_loop.do

This do-file creates the Merged_NoCorp_'year'.dt datasets, which are used in the 5a_MacroAdjustments and 5b_ApplyingMacroAdjustments do files.

Input data:

$databases/Output/AnalysisTaxData/TaxData_Gpercentiles_NoCorp_'year'.dta

$databases/Output/Auxiliary/expansion_gpercentile_NoCorp.dta

$databases/Output/Auxiliary/components_adjustment_gpercentiles_NoCorp.dta

$databases/Output/AnalysisSurveyData/HHSurvey_analysis_'year'.dta

Output data:

$databases/Output/MergedTaxSurvey/Merged_NoCorp_'year'.dta

5b_ApplyingMacroAdjustments.do

This do-file creates the Merged_macro_'year'.dta and Merged_macro_NoCorp_'year'.dta datasets, which are used input files in the 07a_Inequality_YearAnalysis_MacroDistributions and 6a_Inequality_MeasuringIneq do files.

Input data:

$databases/Output/MergedTaxSurvey/Merged_'year'.dta

$databases/Output/Auxiliary/MacroAdjustmentFactors_WithCorp.dta

$databases/Output/Auxiliary/MacroAdjustmentFactors_NoCorp.dta

$databases/AnalysisSurveyData/HHSurvey_analysis_'year'.dta

Output data:

$databases/Output/MergedTaxSurvey/Merged_macro_'year'.dta

$databases/Output/MergedTaxSurvey/Merged_macro_NoCorp_'year'.dta

---

6a_Inequality_MeasuringIneq.do

This do-file creates the InequalityMeasures.dta dataset, which is later used as input in the 6b_Inequality_MeasuringIneq_Graphs do file.

Input data:

$databases/Output/MergedTaxSurvey/Merged_macro_NoCorp_'year'.dta

Output data:

$databases/Output/MergedTaxSurvey/InequalityMeasures.dta

7.2   Analysis

11_CreateTableCounts.do

This do-file creates Table 1, which presents counts of observations in tax records, by year and source of income information.

Input files:

$databases/Output/AnalysisTaxData/TaxData_Analysis_'year'.dta

Output files:

$tables/numbers.tex

3a_Inequality_MergeTaxSurvey_loop.do

This do-file creates Table 2, summarizes the merging process between household survey microdata and administrative tax records.

Input files:

$databases/Output/AnalysisTaxData/TaxData_Gpercentiles_'year'.dta

$databases/Output/Auxiliary/expansion_gpercentile.dta

$databases/Output/Auxiliary/components_adjustment_gpercentiles.dta

$databases/Output/AnalysisSurveyData/HHSurvey_analysis_'year'.dta

Output files:

$databases/Output/MergedTaxSurvey/Merged_'year'.dta

$databases/Output/Auxiliary/merging_diagnostic.dta

$tables/table_sample_paper.tex

Appendix_TaxRecords_Extra.do

This do-file creates Table 3, Table E1, Table E2, Figure 6, Figure 9, Figure 10, Figure 11, Figure A5, Figure E2, and Figure E3. Table 3 documents differences in income composition at top 1% by gender. Table E1 presents the thresholds to belong to each top quantile (in USD 1,000s). Similarly, Table E2 presents the thresholds to belong to each top quantile in local currency. Figure 6 displays the presence of large shareholders at the top of the distribution. Figure 9 shows the median rank in 2014 and 2018 for each bin of income ranking in 2011. Figure 10 presents the share of female taxpayers in the top quantiles of the distribution, pooled for the 2014-2019 period. Figure 11 presents the share of labor and capital income for each quantile, separately for women and men. Figure A5 presents the income composition in terms of different types of income for 2019. Figure E2 shows the average income in each g-percentile within the top 1% for 2019. Finally, Figure E3 presents the share of total income for each g-percentile, taking the average across the 2014-2019 period.

Input files:

    $databases/Output/AnalysisTaxData/TaxData_Analysis_'year'.dta

    $databases/Inputs/AuxData/inflation.dta

    $databases/Output/Auxiliary/survey_population.dta

    $databases/Inputs/AuxData/WID_PPP.dta

Output files:

    $figures/top_distribution_19_nocorp.png

    $figures/top_share_female.png

    $figures/share_capital_income_pooled.png

    $figures/share_income_wid_pooled.png

    $figures/share_capital_income_pooled_undprofit_detailed.png

    $figures/share_income_wid_pooled_undprofit.png

    $tables/thresholds_top.png

    $tables/thresholds_top_lemp.png

    $tables/regression_gender.png

    $figures/share_income_wid_pooled_gender.png

    $figures/share_shareholder_before_after.png

    $figures/share_belowMW.png

5a_MacroAdjustments.do

This do-file creates Figure 1, which presents the adjustment factors applied to the combined household survey + tax record microdata, and Table C2, which documents aggregate variables from the final microdata.

Input files:

    $databases/Output/MergedTaxSurvey/Merged_'year'.dta

    $databases/Output/MergedTaxSurvey/Merged_NoCorp_'year'.dta

    $databases/Output/Auxiliary/MacroAggregatesWDI.dta

Output files:

    $tables/merged_aggregates.tex

    $figures/macro_adjustment_factors_NNI.png

$figures/macro_adjustment_factors.png
$databases/Output/Auxiliary/MacroAdjustmentFactors_WithCorp.dta
$databases/Output/Auxiliary/MacroAdjustmentFactors_NoCorp.dta

---

07a_Inequality_YearAnalysis_MacroDistributions.do

This do-file creates Figure 2, Figure 3, Figure A2, and Table A2. Figure 2 presents the average income for each quantile of the distribution after adjusting to national account aggregates in 2019, both in local currency (Lempiras) as well as USD PPP. Figure 3 presents the composition of total income for each quantile of the distribution after adjusting to national account aggregates. Figure A2 shows the labor and capital income shares evolution over time (2011-2019). Table A2 presents income levels for selected quantiles in different years.

Input files:

$databases/Output/MergedTaxSurvey/Merged_macro_'year'.dta
$databases/Inputs/AuxData/inflation.dta
$databases/Output/Auxiliary/survey_population.dta
$databases/Inputs/AuxData/WID_PPP.dta

Output files:

$figures/top_distribution_macro19.png
$figures/composition_wid_macro_20112019.png
$figures/composition_detail_macro_20112019.png
$figures/labor_capital_overtime.png
$figures/labor_capital_top1.png
$tables/thresholds_top_lemp_usd.tex

7c_wid_TableComparisons.do

This do-file creates Table A3, Table A4, and Table A5. Table A3 displays top income shares for a series of countries. Table A4 documents the average income for the top 10%, top 1% and top 0.01% in the same series of countries. Table A5 provides an overview of the data quality and imputation methods used in each of the countries displayed in the previous tables.

Input files:

$databases/WID/231121_aptinc.xls
$databases/WID/231121_xlcusp.xls
$databases/WID/231121_sptinc.xls

Output files:

$tables/wid-2019.xls

6b_Inequality_MeasuringIneq_Graphs.do

This do-file creates Figure 4 and Figure 5. Figure 4 presents the share of total income accruing to the top 1%, 5% and 10% of individuals with highest income, each year. Figure 5 presents the share of

16

total income accruing to the top 1% each year, using different measures.

Input files:

$databases/Output/MergedTaxSurvey/InequalityMeasures.dta

Output files:

$figures/AllMeasuresNA_merge_long.png

$figures/top1_alternative_definitions.png

$figures/top_baseline_nocorp.png

16_NewTopStability.do

This do-file creates Figure 8, which displays the stability at the top of distribution over time, that this, the probability that individuals in the top 1% or top 0.1% of the income distribution in year t will still be in the same top group in year t+1 (panel a) and t+3 (panel b).

Input files:

$databases/Output/AnalysisTaxData/TaxData_Analysis_'year'.dta

Output files:

$figures/stability_overyears_three.png

$figures/stability_overyears_one.png

PIP_Code.do

This do-file creates Figure A1, which displays growth, poverty, and inequality trends for a subsample of Latin American countries.

Input files:

- `pip` package data

Output files:

$figures/poverty_trends.png

$figures/gini_trends.png

$figures/gdp_trends.png

08_TopStability.do

This do-file creates Figure A3, which documents the stability of individuals at the top (given that we observe an individual at the top of the income distribution in a given year, the figure shows the likelihood that he stays at the top over time.

Input files:

$databases/Output/AnalysisTaxData/TaxData_Analysis_'year'.dta

Output files:

$figures/stability_top.png

Aux_MergingDiagnostic.do

This do-file creates Figure A4, which illustrates the tax-to-survey observation ratios across income levels for three years.

Input files:

$databases/Output/Auxiliary/merging_diagnostic.dta

Output files:

$figures/merging_diagnostic.png

1f_MacroAggregate.do

This do-file creates Figure C1, which shows both the gross domestic product and the net national income evolution over time. It also creates the MacroAggregatesWDI.dta dataset, used as intermediate dataset in the 5a_macroadjustments do file.

Input files:

$databases/Macro/wid_all_data_Apr2023/WID_data_HN.csv

Output files:

$figures/growth_GDP_NNI.png
$databases/Output/Auxiliary/MacroAggregatesWDI.dta

Aux_CorporateMacroAggregates.do

This do-file creates Table B1, which presents some key descriptive statistics of the dataset of corporate income tax (CIT) declarations for every year in the period 2003 - 2019.

Input files:

$databases/Output/Auxiliary/CIT_clean.dta
$databases/Inputs/RawTaxData/DMR_113_FirmLevel.csv
$databases/Output/Auxiliary/UndistributedProfits_relation.dta

Output files:

$tables/tables/table_corporate.tex

08bMacroAggregate_Descriptives.do

This do-file creates Table C1 and Figure C2. Table C1 displays national accounts aggregates. Figure C2 presents the composition of macro incomes from WDI data over time.

Input files:

$databases/Output/Auxiliary/MacroAggregatesWDI.dta

Output files:

$figures/composition_income_macro_wdi.png
$tables/macro_aggregates_wid.tex

Aux_WID_DescribeAggregates.do

This do-file creates Figure C3, which displays the corporate profits and property income dynamics of households in Honduras and other Latin American and high-income countries.

Input files:

$databases/WID/231121_wid_aggregates.xls

Output files:

$figures/corp_prof_hnd_lac.png
$figures/corp_prof_hnd_developed.png
$figures/prop_inc_hnd_lac.png
$figures/prop_inc_hnd_developed.png
$figures/sum_corp_hh_hnd_lac.png

20_AggregatesMacro.do

This do-file creates Figure C4, which shows the composition of macro incomes (wages, mixed incomes, and capital income) over time using merged tax and survey data.

Input files:

$databases/Output/MergedTaxSurvey/Merged_'year'.dta

Output files:

$figures/composition_income_mergeddata.png

Appendix_TaxRecords.do

This do-file creates Figure E2, which presents the distribution of total income in the tax records, excluding the imputed undistributed profits, and the share of capital in total income.

Input files:

$databases/Output/AnalysisTaxData/TaxData_Analysis_2019.dta

Output files:

$figures/histogram_taxincome_2019.png
$figures/composition_taxincome_2019.png

2a_AgriculturalIncomeAdjust.do

This do-file creates Table F1, which presents, for each year in the period 2003-2013, the share of observations for which agricultural income is topcoded and the threshold for topcoding.

Input files:

$databases/Inputs/CleanSurveyData/Nueva Hogar'year'_clean_v2.dta
$databases/Inputs/AuxData/inflation.dta

Output files:

$tables/topcoding_count.tex

2f_Aux_ExpansionGperc.do

This do-file creates Table G1, which presents the adjustment factors used to scale total income by g-percentile, in the period 2003-2010. In the figure, we plot the mean (95% CI) ratio between income from PIT filings + undistributed corporate profits and total income, by g-percentile, across years in the period 2011-2019.

Input files:

$databases/Output/AnalysisTaxData/TaxData_Analysis_'year'.dta

$databases/Output/Auxiliary/survey_population.dta

Output files:

$figures/ajd_factor_gpercentile

$databases/Output/Auxiliary/expansion_gpercentile.dta

7.3   Other Codes

IncomeTaxBrackets.do

This do-file provides locals with information on the minimum wage and the thresholds (tax brackets) for each category of the progressive income tax scheme for each year.

Otros ingresos.do

This do-file creates individual non-labor income for 2011-2018.

Labor_income.do

This do-file creates individual labor income for wages and self-employment activities.

7.4   Programs

Programs installed via SSC

Our code utilizes the following commands installed via SSC manually in `folderDeclare.do`:

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

- `ineqdec0:`
  Stephen P. Jenkins, 1999. "INEQDEC0: Stata module to calculate inequality indices with decomposition by subgroup," Statistical Software Components S366007, Boston College Department of Economics.

- `gtools:`
  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, revised 03 Apr 2019.

- `wid:`
  Blanchet, T. (2021). "WID: Stata module to download data from the World Inequality Database (WID. world)".

- `carryforward:`
  David Kantor, 2004. "CARRYFORWARD: Stata module to carry forward previous observations," Statistical Software Components S444902, Boston College Department of Economics.

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

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