Horizontal Equity of Taxation: Citizen Beliefs and Policy Preferences

Code Repository Documentation

Pierre Bachas
World Bank
pbachas@worldbank.org

Christopher Hoy
World Bank
choy@worldbank.org

Anders Jensen
Harvard & NBER
anders_jensen@hks.harvard.edu

Mahvish Shaukat
World Bank
mshaukat@worldbank.org

Contents

1 Overview

The codes in this repository replicate the tables and figures from "Horizontal Equity of Taxation: Citizen Beliefs and Policy Preferences", by Bachas, Hoy, Jensen, and Shaukat. 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 and programs.

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.

2.3 Details on each Data Source

This study is pre-registered on the AEA Social Science Registry (AEARCTR-0009771). The data for the project consists of two different batch of resources:

- The first data set consists on nationwide representative household surveys of 28 developing counties. Most data come from the restricted access World Bank Microdata Library. If a survey was listed without its microdata on the World Bank platform, we directly contacted the country's' statistical agency.
- The second data set consist on two independent experimental surveys collected by the authors:
  - Experiment 1: In-person Survey with 2,609 individuals in Lahore, Pakistan between May and July 2024.
  - Experiment 2: Online experiment with approximately 15,000 respondents in Colombia, India, Indonesia, Nigeria, and the Philippines.

3 Datasets

3.1 Raw Data

3.1.1 Expenditure Household Survey Raw Data

This section describes the raw data files on household survey data of 28 countries in micro-simulations. The data comes from nationally representative income and expenditure surveys. Some of the micro-data are open access and can be obtained from the statistical agency of each country; others are restricted access. Table 1 describes raw data sources that combined were used to build both expenditure and employment variables. Table 2 describes auxiliary data files used to complement raw houshold survey data files in the micro-simulations.

Table 1: Raw datasets household survey data

Country	Survey (Year)	Expenditure Module	Labor Module	Source	Notes	1st Access
Benin	EICVN (2015)	individu.dta, RM11.dta, RMAL.dta, RM12.dta, RM12a.dta, RM13.dta, RM13a.dta, RM14.dta, RM14a.dta, RM15.dta, RM15a.dta, RM16.dta, RM16a.dta, RM17.dta, RM17a.dta, RM18.dta, RM18a.dta, RM19.dta, RM19a.dta, RM20.dta, RM20a.dta, RM21.dta, RM21a.dta	individu.dta, REMP.dta	Stat. Office	Data is not publicly available and was obtained through World Bank internal access.	Oct 2019
Brazil	POF (2009)	t_morador_s.dta, t_domicilio_s.dta, t_despesa_12meses_s.dta, t_despesa_90dias_s.dta, t_despesa_veiculo_s.dta, t_caderneta_despesa_s.dta, t_despesa_individual_s.dta, t_outras_despesas_s.dta, t_servico_doms_s.dta, t_aluguel_estimado_s.dta, BR_POF_crosswalk_TOR_all.csv, BR_POF_crosswalk_product_code.csv, product_codes.dta, BR_POF_crosswalk_COICOP_all_category_codes.dta, all_retailers_codes.dta, BR_POF_crosswalk_TOR_3dig.xlsx	rend_dedu.dta,	Stat. Office	Data is open access, files can be downloaded from the LINK at least with registration.	Apr 2018
Burkina Faso	EICVM (2009)	EICVM_fichier_menage_produit.dta, CQ02.dta, CQ04.dta, CQ05.dta, CQ06.dta, CQ07.dta, CQ08.dta, CQ09.dta, CQ10.dta, CQ11.dta, CQ12.dta, CQ13.dta, CQ14.dta, CQ15.dta, CQ16.dta, CQ17.dta, CQ18.dta, CQ19.dta, CQ20.dta, CQ21.dta, CQ22.dta, CQ23.dta	individu.dta	Stat. Office	Data is not publicly available and was obtained through World Bank internal access.	Dec 2019
Burundi	ECVM (2014)	poverty_final_version_13.dta, CQ03.dta, CQ04.dta, CQ05.dta, CQ06.dta, CQ07.dta, CQ08.dta, CQ09.dta, CQ10.dta, CQ11.dta, CQ12.dta, CQ13.dta, CQ14.dta, CQ15.dta, CQ16.dta, CQ17.dta, CQ18.dta, CQ19.dta, CQ20.dta, CQ21.dta, CQ22.dta, CQ23.dta	poverty_final_version_13.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	Nov 2017
Cameroon	ECAM (2014)	ECAM4IND.dta, ECAM4MEN.dta, ECAM4_Weight_12052015.dta, ECAM4LOG.dta, hh_income.dta, ECAM4S14.dta, ECAM4PRODUIT.dta	ECAM4IND.dta	World Bank	Data requires registration to request and access.	Mar 2016
Chile	EPF (2017)	BASE_PERSONAS_VIII_EPF.dta, BASE_ESTABLECIMIENTOS_VIII_EPF.dta	BASE_PERSONAS_VIII_EPF.dta	Stat. Office	Data is open access, files can be downloaded from the LINK at least with registration.	May 2018
Colombia	ENIG (2007)	Ig_ml_persona.dta, Ig_gsdu_gas_dia.dta, Ig_gssr_gas_sem.dta, Ig_ml_hogar.dta, CO_dwellings_original.dta, Ig_gsdp_gas_dia.dta, Ig_gsmf_compra.dta, Ig_gsmf_serv_pub.dta	Ig_ml_ocupado.dta, Ig_ml_persona.dta, Ig_gsdu_gas_dia.dta, Ig_gssr_gas_sem.dta	Stat. Office	Data is open access, files can be downloaded from the LINK at least with registration.	Feb 2018
Comoros	EDMC (2013)	demo.dta, apcq03 à 21.dta	emplois_fichier_toutes_personnes.dta, demo.dta	Stat. Office	Data is not publicly available and was obtained through World Bank internal access.	Sep 2019
Congo	ECOM (2005)	ecom2005men.dta, ecom2005pivot.dta	ecom2005ind2.dta, ecom2005men.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	Mar 2014
Costa Rica	ENIGH (2014)	ENIGH_2013_BASE_HOGARES.dta, ENIGH_2013_BASE_GASTOS_EXTENDIDA_PUBLICACION	ENIGH_2013_BASE_HOGARES.dta, ENIGH_2013_BASE_PERSONAS.dta	Stat. Office	Data is open access, files can be downloaded from the LINK at least with registration.	Jan 2018
DR Congo	E123 (2005)	Phase1_IND.dta, Phase3_final_depenses.dta	Phase1_IND.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	Jul 2018
Dominican Rep	ENIGH (2007)	02 cuestionario 1_Poblacion.dta, 01 Datos_Hogar.dta, 03 cuestionarios 1-2-3_Gasto.dta	02 cuestionario 1_Poblacion.dta, 01 Datos_Hogar.dta	Stat. Office	Data is not publicly available and was obtained through World Bank internal access.	June 2019
Ecuador	ENIGHUR (2012)	04_enighur11_personas.dta, 01_enighur11_identificacion.dta, 28_enighur11_gdiarios_seccion2.dta, 29_enighur11_gdiarios_seccion3.dta, 30_enighur11_gdiarios_seccion4.dta, 35_enighur11_gmensuales.dta, 36_enighur11_gtrimestrales.dta, 37_enighur11_gsemestrales.dta, 38_enighur11_ganuales.dta, 31_enighur11_gmensuales_partea.dta, 32_enighur11_gmensuales_parteb.dta, 33_enighur11_gmensuales_partec.dta, 34_enighur11_gmensuales_parted.dta	04_enighur11_personas.dta, 01_enighur11_identificacion.dta	World Bank	Data is open access, files can be downloaded from the LINK at least with registration.	Oct 2018
India	NSS (2012)	Household characteristics - Block 3 - Level2 - type2 - 68.dta, Household characteristics - Block 3 - level3 - type2 - 68.dta, Block 4 - Level 4 - Type 2 - 68.dta, Block 5.1-5.2- 6 - Level 5 - type 2 - 68.dta, Block 7 and 8 - Level 6 - Type 2 - 68.dta, Block 9 - Level 7 - type 2 - 68.dta, Block 10 - Level 8 - type 2 - 68.dta, Block 11 - Level 9 - Type 2 - 68.dta	Household characteristics - Block 3 - Level2 - type2 - 68.dta, Household characteristics - Block 3 - level3 - type2 - 68.dta, Demographic and other particulars of household members - Block 4 - Level 4 - Type 2 - 68.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	May 2023
Indonesia	SUSENAS (2019)	kor19rt_revisi1_diseminasi.dta, kor19ind_1_diseminasi.dta, blok41_gab_diseminasi_prov11_32rev.dta, blok41_gab_diseminasi_prov33_65rev.dta, blok41_gab_diseminasi_prov71_94rev.dta, blok42_revisi_diseminasi_prov11_36, blok42_revisi_diseminasi_prov51_94	kor19ind_1_diseminasi.dta, kor19rt_revisi1_diseminasi.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	Feb 2020
Mexico	ENIGH (2014)	ncv_gastohogar_2014_concil_2010_stata.dta, ncv_concentrado_2014_concil_2010_stata.dta, MX_ENIGH_crosswalk_COICOP.dta	tra_trabajos_2014_concil_2010.dta, ncv_poblacion_2014_concil_2010_stata.dta, ncv_concentrado_2014_concil_2010_stata.dta, ncv_vivi_2014_concil_2010_stata.dta,	Stat. Office	Data is open access, files can be downloaded from the LINK at least with registration.	July 2017
Montenegro	HBS (2009)	indiv_labelled.dta, general_info.dta, diary_labelled.dta, p2stan.dta, p3kupovina.dta, diary_labelled.dta, nonfood items.dta	indiv_labelled.dta, general_info.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	Apr 2013

Country	Survey (Year)	Expenditure Module	Labor Module	Source	Notes	1st Access
Morocco	ENCDM (2001)	fichier_individu_encdm2000_01.dta, BM_with_hhid.dta, expenses_and_totals.dta, MA_ENCDM_crosswalk_COICOP.csv, Morocco_codes_1dig.dta, Morocco_codes_2dig.dta, Morocco_codes_3dig.dta, Morocco_codes_4dig.dta	fichier_individu_encdm2000_01.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	May 2013
Mozambique	IOF (2009)	agregado_v3.dta, controle_v3.dta, iof_ponderadores3.dta, despesas_diarias_v3.dta, despesas_mensais_v3.dta, bens_duraveis_v3.dta, autoconsumo_v3.dta, receita_v3.dta, verificacao_de_produtos_v3.dta, agregado_v3.dta, iof_ponderadores3.dta	agregado_v3.dta, iof_ponderadores3.dta, controle_v3.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	May 2013
Niger	ENBCM (2007)	r02.dta, menageidentifiant.dta, r01.dta, r15.dta, enbc2007pivot.dta, menageidentifiant.dta	menageidentifiant.dta, r05a1.dta, r05c.dta, r02.dta, r01.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	May 2013
Pakistan	PSLM (2019)	plist.dta, sec_5a.dta, sec_2ab.dta, sec_6a.dta, sec_12ce.dta, sec_12a.dta	plist.dta, sec_5a.dta, sec_2ab.dta, sec_1b (2).dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	Mar 2020
Paraguay	EIGyCV (2011)	Datos de la poblacion reg02t.dta, Datos de la vivienda y del Hogar reg01.dta, Agregado a nivel de transaccion agregado.dta	Datos de la poblacion reg02t.dta, Datos de la vivienda y del Hogar reg01.dta	World Bank	Data is not publicly available and was obtained through World Bank internal access.	Dec 2020
Peru	ENAHO (2017)	enaho01-2017-200.dta, enaho01a-2017-300.dta, enaho01-2017-100.dta, enaho01-2017-601.dta, enaho01-2017-602.dta, enaho01-2017-603.dta, enaho01-2017-604.dta, enaho01-2017-605.dta, enaho01-2017-606.dta, enaho01-2017-606d.dta, enaho01-2017-607.dta, enaho01-2017-610.dta, enaho01-2017-611.dta, enaho01-2017-612.dta, enaho01a-2017-300.dta, enaho01a-2017-400.dta, enaho01-2017-100.dta, PE_ENAHO_crosswalk_TOR.dta, PE_ENAHO_crosswalk_product_code.xls, PE_COICOP_crosswalk_mod_1.dta, PE_COICOP_crosswalk_mod_3.dta, PE_COICOP_crosswalk_mod_4.dta, PE_COICOP_crosswalk_mod_7.dta, PE_COICOP_crosswalk_mod_8.dta, PE_COICOP_crosswalk_mod_9.dta, PE_COICOP_crosswalk_mod_10.dta, PE_COICOP_crosswalk_mod_11.dta, PE_COICOP_crosswalk_mod_12.dta, PE_COICOP_crosswalk_mod_13.dta, PE_COICOP_crosswalk_mod_16.dta, PE_COICOP_crosswalk_mod_17.dta, PE_COICOP_crosswalk_mod_18.dta, PE_COICOP_crosswalk_mod_78.dta	enaho01-2017-500.dta, enaho01-2017-200.dta, enaho01a-2017-300.dta, enaho01-2017-100.dta	Stat. Office	Data is open access, files can be downloaded from the LINK.	Apr 2018
Phillipines	FIES (2018)	2018_FIES_vol2.csv	2018_FIES_vol2.csv, FIES PUF 2018 Vol.1.csv	World Bank	Data is not publicly available and was obtained through World Bank internal access.	Feb 2023
Rwanda	EICV (2014)	cs_s1_s2_s3_s4_s6a_s6e_s6f_person.dta, cs_s8b_expenditure.dta, cs_s8a3_expenditure.dta, cs_s8a2_expenditure.dta, cs_s8a1_expenditure.dta, cs_s8c_farming.dta, cs_s0_s5_household.dta, cs_s6b_employement_6c_salaried_s6d_business.dta, cs_s9b_transfers_in.dta, cs_s10b_goods.dta	cs_s6b_employement_6c_salaried_s6d_business.dta, cs_s1_s2_s3_s4_s6a_s6e_s6f_person.dta	World Bank	Data is open access, files can be downloaded from the LINK.	jul 2016
Senegal	EDMC (2008)	CQ02Aan.dta, CQ06Aan.dta	CQ06Aan.dta, CQ03Aan.dta, CQ02Aan.dta	World Bank	Data is open access, files can be downloaded from the LINK at least with registration.	May 2019
Tanzania	HBS (2012)	from_SECT1.dta, SECTA1.dta, SECTA2.dta, SECTB1.dta, SECTB3.dta, SECT2.dta, SECT2Q27.dta, SECT4A.dta, SECT5.dta, SECT6Q15.dta, SECT7Q1.dta, SECT7Q9.dta, SECT7Q11.dta, SECT7Q13.dta, SECT7Q15.dta, SECT8Q2.dta, SECT8Q4.dta, SECT8Q6.dta, SECTB1.dta	SECT12.dta, SECT1.dta	World Bank	Data is open access, files can be downloaded from the LINK.	Jul 2018
Tunisia	EBC (2010)	bdg2010T2.dta, produit.dta, menage.dta, produit.dta	menage.dta, bdg2010T2.dta	Stat. Office	Data is not publicly available and was obtained through World Bank internal access.	March 2015

Table 2: Additional data files for micro-simulations

Dataset	Description	Notes	1st Access
countrycode2d_region.csv	Data file icnluding country codes.	Retrieved from github.com/datasets/country-codes	Jun 2023
API_NY.GDP.PCAP.PP.KD_DS2_en_excel_v2_5607693.xls	GDP per capita, PPP (Constan 2021 prices).	Retrieved from World Bank Development Indicators https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD	Jul 2023
API_PA.NUS.PPP_DS2_en_excel_v2_5607721.xls	PPP conversion factor, GDP (LCU per international $).	Retrieved from World Bank Development Indicators https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD	Jul 2023
TaxDevEITDv0.2.xlsx	Data file including detailed description of PIT parameters across countries.	Retrieved from McNabb, K. (2022). *The TaxDev employment income taxes dataset: Technical guide v1.0.* https://odi.org/en/publications/the-taxdev-employment-income-taxes-dataset-and-technical-guide and public government records.	Nov 2024
isocodes_dictionary.dta	Cross walk relating 2 and 3 digits ISO codes across countries.	Generated in `1a_auxiliar_data.do`	Nov 2024
saving_rates.dta	Average saving rates across income deciles at the country level.	Retrieved from *Global Findex Database (Global Financial Inclusion).* Washington, DC: World Bank https://datacatalog.worldbank.org/int/search/dataset/0039935	Jul 2024
VAT_DATA.xlsx	Data file including detailed description of VAT parameters across countries.	Retrieved from Worldwide Tax Summaries PwC (2025). *Worldwide Tax Summaries.* https://taxsummaries.pwc.com/quick-charts/value-added-tax-vat-rates and complemented with official government data sources for each country.	Jun 2022
wb_stats.dta	Data file including relevant variables across countries.	Retrieved from World Bank Open Indicators https://databank.worldbank.org/source/world-development-indicators.	Feb 2025
gethin-public-finance-statistics.dta	Data file with extensive statistics on government income and expenditure at the country-year level.	Retrieved from https://www.dropbox.com/scl/fi/i1wo4bzx1eg9ucbnzjfu5/gethin-public-finance-statistics-2025-01-21.dta?rlkey=7qy5dfwg3fqvwty3shx9wd6fc&e=1&st=mgo9q3hx&dl=0. For details see Gethin (2024), *A New Database of General Government Revenue and Expenditure by Function, 1980-2022.*	Nov 2024
dataset_formalityshare_actualpredict.dta	Data file with country specific probabilities for modern and traditional using Euromonitor retail reports.	Data file was obtained from Bachas et al (2023) https://doi.org/10.1093/restud/rdad095.	Nov 2024
country_list_for_loop.xlsx	Meta data file for the list of countries in microsimulations.	The source is own elaboration. Includes the name of the countries in the microsimulations along with the names of all relevant files required.	Jul 2025
world-imf2024-v1.dta	Data file at the country-year level with tax collection.	Data source of tax collection is World Revenue Longitudinal Database from the International Monetary Fund https://www.imf.org/en/topics/fiscal-policies/world-revenue-longitudinal-database.	Dec 2025
OGHIST_2025_10_07.xlsx	Data file with income level classification from the World Bank.	Data retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups	Jan 2026
COICOP_label_2dig.dta COICOP_label_3dig.dta COICOP_label_4dig.dta	Cross walk data files with Classification of Individual Consumption According to Purpose (COICOP) at the 2, 3, and 4 digits level.	Data retrieved from https://unstats.un.org/unsd/classifications/coicop.	Jun 2023
Country_information.xlsx	Data file including income per capita and conversion rates.	Data file compiled by hand from different data sources. GDP data at: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD. Gini data at: https://data.worldbank.org/indicator/SI.POV.GINI/. Conversion factor at: https://data.worldbank.org/indicator/PA.NUS.PPP	Mar 2021
wbopendata_metadata.dta	Data file including country metadata as reported in the World Bank Development Indicators.	Run `wbopendata` comand in Stata and recover metadata. Alternative is recovering metadata directly from https://databank.worldbank.org/source/world-development-indicators	Mar 2026
OECD.SDD.NAD,DSD_NAMAIN10@DF_TABLE1_EXPENDITURE,2.0,filtered,2026-02-19 18-17-16.xlsx OECD.SDD.NAD,DSD_NAMAIN10@DF_TABLE1_EXPENDITURE,2.0,filtered,2026-02-19 18-27-01.xlsx	Data files from OECD National Accounts on final consumption expenditure for households and NPISH.	Retrieved from https://data-explorer.oecd.org/, Annual GDP and components - expenditure approach.	Feb 2026
HH and NPISHs FCE - From 2000.xlsx	Data file from World Bank Development Indicators with household and NPISH final consumption expenditure from 2000 onwards.	This data is not publicly available and was shared by DECRG Data Team of the World Bank. Please contact: data@worldbank.org	Feb 2026
WID_Data_27022026-175951.xlsx	Data file with pre-tax income distribution at the percentile level across countries.	Retrieved from World Inequality Database https://wid.world/.	Feb 2026
Bachas_coef_allcountries.dta	Data file with coefficients from cross-country regressions on informality consumption shares.	Coefficientes were retrieved from column (5) Table from supplementary online Appendix of Bachas et al (2023) https://doi.org/10.1093/restud/rdad095. The results come from a regression where the dependent variable is the informal share of households' total consumption and the covariates are the logarithm of households' expenditure (income proxy), household size, household head's age, gender of household head gender (dummy if female), and a dummy for food items.	Mar 2026
country_validation.dta	Data file including aggregate outcomes at the country level.	Data manually gathered from World Bank Open Indicators https://databank.worldbank.org/source/world-development-indicators.	Aug 2025

3.1.2 Experimental Surveys Raw Data

This sections describes raw data files from two survey experiments conducted by the authors. The first dataset comes from an in-person survey experiment carried out in Lahore, Pakistan between May and July 2024. The survey sampled 2,609 workers.

The second dataset comes from an online survey experiment conducted in Colombia, India, Indonesia, Nigeria, and the Philippines. The survey was conducted with approximately 15,000 respondents, representative of internet users in each country. Respondents were randomly assigned to one of four groups: an information treatment that provided data on the share of taxes paid by self-employed workers compared to their representation in the workforce; a pedagogical treatment that explained how consumption taxes contribute to greater horizontal equity; a combined treatment that included both; or a control group that received no information. Table 3 describes raw data sources that combined were used to build both expenditure and employment variables.

Table 3: Raw datasets experimental surveys

Dataset	Description	Raw data files	Notes	1st Access
In-person survey instrument (Lahore, 2024)	In-person survey experiment in Lahore, Pakistan.	`240712_Horizontal_Equity_Survey_Rollout_*.dta` `stratavar.dta`		July 2024

Online survey instrument (Multi-country, 2024)	Online survey experiment in Colombia, India, Indonesia, Nigeria, and the Philippines.	`raw.dta`
		`wb_tax_compliance_2_20220824.xlsx`
		`wb_tax_compliance_2_20220824.sav`
			Exclude folder: Files in `/exclude` were excluded from the raw data but appended in `cleaning.do`. Available in both `.dta` and `.csv` formats:
			`data/raw/exclude/excluded_CO.dta`
			`data/raw/exclude/excluded_IN.dta`
			`data/raw/exclude/excluded_ID.dta`
			`data/raw/exclude/excluded_NG.dta`
			`data/raw/exclude/excluded_PH.dta`
				July 2024

3.2 Processed Data

This section describes the processed data obtained after cleaning raw files in both the micro-simulations and experimental survey analysis. The last column "`Provided`" in each of the tables indicates wether the data file in mention is publicly available along with this replication package or not.

3.2.1 Expenditure Household Survey Processed Data

Table 4: Processed data files for micro-simulations

Dataset	Description	Notes	Provided
ppp conversion factor.dta	Data file with power purchasing parity conversion factors at the country-year level.	Confidential	No
country income per capita ppp.dta	Data file with income per capita at the country level.	Confidential	No
*_labor_head.dta	Harmonized data files with labor modules at the country level. In each file, the suffix * denotes the ISO code of each country.	Confidential	No
*_household_cov_original.dta *_household_cov.dta *_all_lines_raw.dta *_TOR_stats_for_crosswalk.xls *_exp_full_db.dta *_temp_exp_TOR_item_COICOP_4dig.dta	Harmonized data files with labor modules at the country level. In each file, the suffix * denotes the ISO code of each country.	Confidential	No
all countries_income informality.dta	Data file with variables on informality consumption for a selected sample of countries and is used to infer the share of informality consumption in Pakistan.	Confidential	No
*_VAT simulation data.dta	Harmonized data files including expenditure relevant outcomes adjusted by informality patterns. In each file, the suffix * denotes the ISO code of each country.	Confidential	No
cross_country_hh_total_expenditure.dta	Data file including aggregate expenditure at the country level as in household surveys.	Confidential	No
consumption_national_accounts.dta	Data file including household final consumption expenditure and NPISH consumption from national accounts, adjusted to be comparable with household surveys.	Confidential	No
median_income_wid.dta	Data file including median pre-tax income and PIT exemption threshold at the country level, used to match imputed income in simulations.	Confidential	No
consumption_adjustment_factor.dta	Data file including adjustment factor at the country level for the ratio between aggregate expenditure in household surveys vs aggregate consumption in national accounts.	Confidential	No
final_metadata_exp*_se*_emp*_adj*.dta	Data file including key tax parameters—the exemption threshold (first bracket bound), the top marginal tax rate—for a panel of countries and share of employment above the exemption threshold. Each * represents the global parameters used in the micro-simulations.	Confidential	No
${country_code}_${year}_emp.dta	Data file including the employment shares for each decile at the country level.	Confidential	No
final_employment_meta_exp*_se*_emp*_adj*.dta	Stacked cross-country version of ${country_code}_${year}_emp.dta. Each * represents the global parameters used in the micro-simulations.	Confidential	No
${country_code}_${year}_cons.dta	Data file including consumption patterns for each decile at the country level.	Confidential	No
final_consumption_meta_exp*_se*_emp*_adj*.dta	Stacked cross-country version of ${country_code}_${year}_cons.dta	Confidential	No
*_brackets_top_earners.dta	Data file including the brackets of the tax rates paid by top income households. In each file, the suffix * denotes the ISO code of each country.	Confidential	No
*_linear rev_pit.dta	Data file with estimates of the slope for relationship between PIT revenue and changes in marginal tax rates. In each file, the suffix * denotes the ISO code of each country.	Confidential	No
*_linear rev_vat.dta	Data file with estimates of the slope for relationship between VAT revenue and changes in the statutory rate. In each file, the suffix * denotes the ISO code of each country.	Confidential	No
${country_code}_sim_reform_exp*_se*_emp*_adj*.dta	Data files including micro-simulation results for tax burden and country specific estimates of horizontal and vertical equity measures. In each file, the suffix * de- notes the ISO code of each country, and each * represents the global parameters used in the micro-simulations.	Confidential	No
crosscountry_sim_reform_exp*_se*_emp*_adj*.dta	Stacked cross-country version of ${country_code}_sim_reform_exp*_se*_emp*_adj*.dta	Confidential	No
crosscountry_top_brackets_sim_reform_exp*_se*_emp*_adj*.dta	Stacked cross-country version of *_brackets_top_earners.dta	Confidential	No

3.2.2 Experimental Surveys Processed Data: Online Survey

Table 5: Processed datasets: online survey experiment

Dataset	Description (files)	Notes	Provided
Data/Clean	`Clean.dta` (master clean dataset; full sample prior to exclusions); `Analysis_sample.dta` (final analysis sample after exclusions); `cleanforqvals.dta` (cleaned dataset used for q-value regressions)	Final, analysis-ready	No
Data/Intermediate	`Clean_trim_by_ctry_treat.dta` (trimmed within country×treatment at 5th–95th percentile of `duration`); `Intermediate.dta` (intermediate merge/partial cleaning snapshot)	Work-in-progress	No

3.2.3 Experimental Surveys Processed Data: In-person Survey

Table 6: Processed datasets: in-person survey experiment

Dataset	Description (files)	Notes	Provided
Data/Clean	`cleaned_data.dta` (final cleaned dataset from the in-person survey); `cleaned_online.dta` (cleaned dataset from the online experiment conducted in five additional countries); `Clean-multitesting.dta` (dataset used for multiple-hypothesis testing, created from `3_regressions_overleaf.do`)	Final, analysis-ready	No

4 Computational Requirements

4.1 Software requirements

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

`R` is needed to conduct analysis for one figure (Figure B.7).

4.2 Memory and runtime requirements

The first part of the code on micro-simulations 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 143 minutes. The cleaning process alone takes 54 minutes.

The second part of the code on the in-person survey and online experiment was run for the last time on a laptop running Windows 11 Enterprise (OS build 26200.7462), with 32.0 GB RAM (31.5 GB usable), and an Intel(R) Core(TM) Ultra 7 165U processor (2.10 GHz). With those settings, a replicator could expect the whole code to run in 15 minutes.

5 Instructions to replicators

This section shows the structure of every block of code that replicates the results in the paper. The codes described in section 7 rely on the folder structure for each block of the paper as described next.

Important: Each section of the paper must be reproduced independently. This is why the replication package includes two folders: `Reproducibility_microsimulations` for micro-simulations, and `Reproducibility_experiments` for the two field experiments.

5.1 Micro-Simulations: Folder Structure

For full replication of micro-simulations, please refer to folder called `Reproducibility_microsimulations`, which has the following structure:

`1_raw_data`
- `1.1_surveys`

`2_proc`

`3_waste`

`4_dofiles`

`5_graphs`

`6_tables`

Before replication Before running the code and to make the replication process more efficient, in line 41 of section "Set working paths" in the master do-file `0a_master_horizontal_equity.do` the reviewer must insert the user name and the location where raw data and output files are located.

5.2 In-Person and Online Survey Experiments: Folder Structure

This section describes the folder structure used for the in-person and online survey experiments and the associated code required to replicate the results presented in the paper. Please refer to the folder called `Reproducibility_experiments`.

5.2.1 Online Survey

The online survey replication materials can be found in `Reproducibility_experiments/online survey` and are organized into data, code, and output folders as follows:

- `1_data`
- `2_do` — Stata do-files used to clean the data, construct variables, and run regressions.
- `3_output` — Automatically generated outputs:
  - `1_table` — Regression and summary tables.
  - `2_img` — Exported figures in image formats.
  - `3_gph` — Stata graph files.

Before replication Before running the code and to make the replication process more efficient, in line 37 of section "`Set working paths`" in the master do-file the reviewer must insert the user name and the location where raw data and output files are located.

5.2.2 In-Person Survey

The in-person survey replication materials can be found in `Reproducibility_experiments/inperson survey` and follow an analogous structure:

- `1_data`
- `2_do` — Stata do-files used for data cleaning, variable construction, and estimation.
- `3_output` — Automatically generated outputs:
  - `1_table` — Regression and summary tables.
  - `2_img` — Exported figures in image formats.
  - `3_pdf` — Exported figures in pdf formats.

Before replication Before running the code and to make the replication process more efficient, in line 30 of section "`Set working paths`" in the master do-file the reviewer must insert the user name and the location where raw data and output files are located.

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

Table 7: Mapping of tables and analysis code

Exhibit name	**Output filename**	**Code**
Table 1	`table1.tex`	`3_regressions_overleaf.do`*
Table 2	`table2.tex`	`3_regressions_overleaf.do`*
Table 3	`table3.tex`	`3_regressions_overleaf.do`*
Table 4	`table4.tex`	`3_regressions_overleaf.do`*
Table 5	`table5.tex`	`3_regressions_overleaf.do`*
Table A.1	`cross_country_sample.tex`	`4a_tables.do`**
Table A.2	`targeted_moments_exp2_se50_emp100_adj1.tex`	`4a_tables.do`**
Table A.3	`he_ve_estimates_exp2_se50_emp100_adj1.tex`	`4a_tables.do`**
Table A.4	`he_ve_robustness_exp2_se50_emp100_adj1.tex`	`4a_tables.do`**
Table B.1	`TableB1.tex`	`4_final_descriptives.do`*
Table B.2	`TableB2.tex`	`4_final_descriptives.do`*
Table B.3	`B3_table1_nocontrols.tex`	`3_regressions_overleaf_nocontrols.do`*
Table B.4	`B4_table2_nocontrols.tex`	`3_regressions_overleaf_nocontrols.do`*
Table B.5	`B5_table3_nocontrols.tex`	`3_regressions_overleaf_nocontrols.do`*
Table B.6	`B6_table4_nocontrols.tex`	`3_regressions_overleaf_nocontrols.do`*
Table B.7	`multiplehypothesistesting.txt`	`5_multiple_hypothesis_updated.do`*
Table B.8	`multiplehypothesistesting.txt`	`5_multiple_hypothesis_updated.do`*
Table B.9	`multiplehypothesistesting.txt`	`5_multiple_hypothesis_updated.do`*
Table B.10	`multiplehypothesistesting.txt`	`5_multiple_hypothesis_updated.do`*

Table D.1	`balance_tp_all_arms.tex`
	`balance_tp_all_arms_colombia.tex`
	`balance_tp_all_arms_india.tex`
	`balance_tp_all_arms_nigeria.tex`
	`balance_tp_all_arms_philippines.tex`
	`balance_tp_all_arms_indonesia.tex`
		`5_tables.do (online)`*

Table D.2	`D2_online.tex`	`5_tables.do (online)`*
Table D.3	`D3_table1withdem.tex`	`3_regressions.do (online)`*

* Do file from the `Reproducibility_experiments` package. ** Do file from the `Reproducibility_microsimulations` package. P-values for tables B7-B10 coming from Multiple Hypothesis Testing are located in the pdf (multiplehypothesistesting.txt). The remaining values reported in tables B7-B10 (coefficients, number of observations, and control means) are obtained from and thus can be verified in main results (Tables 1-4).

6.2 Mapping of figures and analysis code

Table 8 provides a mapping between all figures of the paper (including the online appendix) and the codes producing these results. The codes generate files in .pdf and .png format containing the graphs.

Table 8: Mapping of figures and analysis code

Exhibit Name	**Panels**	**Output filename**	**Code**
Figure 1	a, b	`share_se_p90_cross_count_exp2_se50_emp100_adj1.pdf`, `share_se_pitexmp_cross_count_exp2_se50_emp100_adj1.pdf`	`4b_figures.do**`
Figure 2	a, b	`shares_inf_s_se_ci_exp2_se50_emp100_adj1.pdf`, `eff_VAT_topdecile_exp2_se50_emp100_adj1.pdf`	`4b_figures.do**`
Figure 3	a, b	`he_ve_scatter_exp2_se50_emp100_adj1_v2.pdf`, `he_ve_arrows_exp2_se50_emp100_adj1_v2.pdf`	`4b_figures.do**`
Figure 4		`Info_treatment_part1_cropped.pdf`, `Info_treatment_part2_cropped.pdf`	Own elaboration
Figure 5	a-c	`Fig8_PanelA_combined.png`, `Fig8_PanelA_combined.png`, `Fig8_PanelA_combined.png`	`4_final_descriptives.do*`
Figure A.1	a, b	`cross_country_vat_pit_revenue.pdf`, `cross_country_vat_pit_gdp_ratio.pdf`	`4b_figures.do**`
Figure A.2	a-c	`topmtr_cross_count_exp2_se50_emp100_adj1.pdf`, `exp_thres_cross_count_exp2_se50_emp100_adj1.pdf`, `etr_above_1_cross_count_exp2_se50_emp100_adj1.pdf`, `etr_above_5_cross_count_exp2_se50_emp100_adj1.pdf`	`4b_figures.do**`
Figure A.3		`PK_ETR_exp2_se50_emp100_adj1.png`	`3a_micro_simulations_baseline.do**`
Figure A.4		`composition_imputed_income_exp2_se50_emp100_adj1.pdf`	`4b_figures.do**`
Figure A.5		`vat_rates_exp2_se50_emp100_adj1.pdf`	`4b_figures.do**`
Figure A.6		`he_ve_scatter_income_robustness.pdf`, `he_ve_arrows_income_robustness.pdf`	`4b_figures.do**`
Figure B.7		`variable_importance_two_panel.pdf`	`6_causal forest figure.rtf*`
Figure D.1		Screenshot from the survey platform	
Figure D.2		Screenshot from the survey platform	
Figure D.3		Screenshot from the survey platform	
Figure D.4	a, b, c	`D4_UNFAIRNESS.png`	`4_figures.do*`
Figure D.5	a, b, c	`D5_POLICY.png`	`4_figures.do*`

* Do file from the `Reproducibility_experiments` package. ** Do file from the `Reproducibility_microsimulations` package.

7 Description of programs and code

The project is organized as follows. First, we described the codes implemented for analysis and micro-simulations using household survey data in section 7.1. Second, we describe the code for the analysis of experiment in section 7.2. Some of the code relies on programs downloaded from the Statistical Software Components (SSC) archive. These programs are described in section 7.3.

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

7.1 Codes for Micro-Simulations

- `0b_install_packages.do`
  Install all packages required to reproduce the code. Automatically checks if packages are already installed before attempting installation to avoid redundant operations. Installs packages from SSC, and user-written packages from external sources (latexify, xfill).

  Inputs:
  None

  Outputs:
  None (packages installed in Stata environment)

- `0c_programs.do`
  Create programs for repetitive tasks needed throughout the code. Defines country-specific deduction programs for PIT simulations (`deductions_XX` for each country), a program to calculate PIT liability based on imputed income following Algorithm 1, a program to calculate VAT liability accounting for informality and differentiated goods following Algorithm 2, country-specific programs to compute gross entry thresholds for the PIT (`gross_entry_threshold_XX`), country-specific deduction programs applied to imputed income (`deductions_imp_XX`), and a program to convert micro-simulation outcomes from household to individual level.

  Inputs:
  None

  Outputs:
  `deductions_XX` (where XX is the country ISO code)
  `calculate_pit_liability`
  `calculate_vat_liability`
  `gross_entry_threshold_XX` (where XX is the country ISO code)
  `deductions_imp_XX` (where XX is the country ISO code)
  `convert_household_to_individual`

- `0d_globals_parameters.do`
  Set global parameters and macros required to run the code and replicate results. Defines simulation parameters (misreporting rates, income definitions, adjustment factors, imputed rent share),

graphing settings, and creates global macros for country loops by importing country list and extracting country codes, names, file names, and other country-specific variables. Also sets the scenario flag to distinguish baseline from robustness runs.

Inputs:

$main/data/country_list_for_loop.xlsx

$main/data/Bachas_coef_allcountries.dta

Outputs:

Global macros: `nocountries`, `country_codes`, `country_fullnames`, `exp_file_names`, `income_measures`, `product_codes`, `categories_4`, `slopes_4`, `bgj_list`, `years`, `mean_income_ratio_usd`, `pit_emp_compl`, `pit_se_compl`, `whichexp`, `adj_factor`, `scenario`, `se_compl`, `emp_compl`, `graphlook`, `dofiles`, `run_empl_list`, `run_etr_list`, `imputed_rent`, `microsim_scenario`, `coef_bachas_income`, `coef_bachas_cons`, `coef_bachas_hhsize`, `coef_bachas_head_sex`, `coef_bachas_age`, `coef_bachas_food`

- `1a_auxiliar_data.do`

  Create auxiliar data files with GDP per capita, or PPP conversion factos across countries.

  Inputs:

  $countrycode2d_region.csv

  $main/data/API_NY.GDP.PCAP.PP.KD_DS2_en_excel_v2_5607670.xls

  $main/data/API_PA.NUS.PPP_DS2_en_excel_v2_5607721.xls

  Outputs:

  $main/data/country income per capita ppp.dta

  $main/data/ppp conversion factor.dta

- `1b_cleaning_hh_surveys.do`

  Run a loop for each country in the sample to clean both labor and expenditure modules in household survey data. Then, produce harmonized final working data files ready for analysis. In total this do files runs 50 additional do files to clean household expenditure data. Appendix A provides detailed description of each do file.

  Inputs:

  Labor data take as input each do file in folder $main/dofiles/employment

  Expenditure data take as input each do file in folder $main/dofiles/expenditure

  Outputs:

  Harmonized labor data files named:

  $survey/proc/$country_fullname/*_labor_head.dta

  Harmonized expenditure data files named:

  $survey/proc/$country_fullname/*_household_cov_original.dta

  $survey/proc/$country_fullname/*_household_cov.dta

  $survey/proc/$country_fullname/*_all_lines_raw.dta

  $survey/proc/$country_fullname/*_TOR_stats_for_crosswalk.xls

  $survey/proc/$country_fullname/*_exp_full_db.dta

  $survey/proc/$country_fullname/*_temp_exp_TOR_item_COICOP_4dig.dta

  Where the suffix * denotes the ISO code of each country

- 1c_cleaning_hh_surveys_prepare_vat_bgj.do

  Run a loop for each country in the sample that has a detailed expenditure module (i.e., the `bgj` flag equals 1) and harmonize expenditure outcomes by adjusting consumption on differentiated goods by informality levels, as in Bachas et al (RES 2023). Country-specific informality weights are calibrated using country-specific formality shares from Euromonitor retail reports when available, with gradient adjustments across expenditure deciles when applicable.

  Inputs:

  $main/data/dataset_formalityshare_actualpredict.dta
  $main/data/Country_information.xlsx
  $survey/proc/${country_fullname}/${country_code}_${exp_file_name}.dta
  $survey/proc/${country_fullname}/${country_code}_labor_head.dta

  Outputs:

  $main/waste/*_VAT simulation data.dta

  Where the suffix * denotes the name of each country survey

- 1c_cleaning_hh_surveys_prepare_vat_nonbgj.do

  Run a loop for each country in the sample that does *not* have a detailed expenditure module (i.e., the `bgj` flag equals 0) and predict informality shares using regression coefficients estimated from the subsample of countries with detailed modules. At the end it produces harmonized expenditure datasets adjusting consumption on differentiated goods by predicted informality levels, following the same approach as Bachas et al (RES 2023).

  Inputs:

  $main/waste/Benin2015_VAT simulation data.dta
  $main/waste/${country_loop_name}_VAT simulation data.dta (stacked BGJ countries)
  $main/data/country income per capita ppp.dta
  $main/data/ppp conversion factor.dta
  $survey/proc/${country_fullname}/${country_code}_temp_exp.dta
  $survey/proc/${country_fullname}/${country_code}_labor_head.dta

  Outputs:

  $main/waste/${country_fullname}_VAT simulation data.dta

- 1d_cleaning_consumption_national_accounts.do

  Run a loop for each country in the sample to get total household expenditure and compare this total to that reported in national accounts; the result is a ratio, at the country level, which is used afterwards as an adjustment factor. In addition, this file retrieves national accounts data from OECD and World Bank sources, predicts NPISH and FISIM consumption for countries with missing data, and retrieves median pre-tax income from the World Inequality Database (WID) to use as a moment for matching imputed income in the micro-simulations.

  Inputs:

  $${country_code}_${exp_file_name}.dta

$${country_code}_labor_head.dta
$main/data/OECD.SDD.NAD,DSD_NAMAIN10@DF_TABLE1_EXPENDITURE,2.0,filtered,2026-02-
$main/data/OECD.SDD.NAD,DSD_NAMAIN10@DF_TABLE1_EXPENDITURE,2.0,filtered,2026-02-
$main/data/HH and NPISHs FCE - From 2000.xlsx
$main/data/WID_Data_27022026-175951.xlsx
$main/data/TaxDevEITDv0.2.xlsx
$main/data/isocodes_dictionary.dta
$main/data/wbopendata_metadata.dta

Where ${country_code} is the ISO code of each country, and ${exp_file_name} is the name of the file with household expenditures for that country.

Outputs:

$main/proc/cross_country_hh_total_expenditure.dta
$main/waste/consumption_national_accounts.dta
$main/waste/median_income_wid.dta
$main/proc/consumption_adjustment_factor.dta

- 2a_cross_country_tax_systems.do

  Automates the extraction of key tax parameters—the exemption threshold (first bracket bound), the top marginal tax rate—for a panel of countries and share of employment above the exemption threshold, merges them with macroeconomic indicators, and produces final data file.

  Inputs:

  $survey/proc/${country_fullname}/*_${exp_file_name}.dta
  $survey/proc/${country_fullname}/*_labor_head.dta
  $main/data/TaxDevEITDv0.2.xlsx
  $main/data/isocodes_dictionary.dta
  $main/data/saving_rates.dta
  $main/data/VAT_DATA.xlsx
  $main/Notes/country_validation.dta
  $main/data/wb_stats.dta
  $main/proc/consumption_adjustment_factor.dta

  Where the suffix * denotes the ISO code of each country

  Outputs:

  $main/proc/final_metadata_exp*_se*_emp*_adj*.dta

- 2b_cross_country_shares_empl_selfempl.do

  This file generates country-level meta data of the share of employed vs self-employed across deciles for each country in the sample and produces final data file stacking all of them.

  Inputs:

  $survey/proc/${country_fullname}/*_${exp_file_name}.dta
  $survey/proc/${country_fullname}/*_labor_head.dta
  $main/data/TaxDevEITDv0.2.xlsx

$main/proc/consumption_adjustment_factor.dta
$main/Notes/country_validation.dta
$main/data/wb_stats.dta
$main/data/gethin-public-finance-statistics.dta
$main/data/isocodes_dictionary.dta
$main/data/saving_rates.dta

Where the suffix * denotes the ISO code of each country

Outputs:

$main/proc/emp_meta/${country_code}_${year}.dta
$main/proc/final_employment_meta_exp*_se*_emp*_adj*.dta

- 2c_cross_country_consumption_empl_selfempl_bgj.do

  This file generates country-level data on the consumption patterns of both employees and self-employees across deciles for each country in the sample and produces final data file stacking all of them. For countries without detailed expenditure modules (`bgj` flag equals 0), it internally calls `2c_cross_country_consumption_empl_selfempl_nonbgj.do` to handle the alternative computation approach.

  Inputs:

  $survey/proc/${country_fullname}/*_${exp_file_name}.dta
  $survey/proc/${country_fullname}/*_labor_head.dta
  $main/data/dataset_formalityshare_actualpredict.dta
  $survey/proc/${country_fullname}/*_household_cov.dta
  $main/data/wb_stats.dta
  $main/data/gethin-public-finance-statistics.dta

  Where the suffix * denotes the ISO code of each country

  Ouputs:

  $main/proc/cons_meta/${country_code}_${year}_cons.dta
  $main/proc/final_consumption_meta_exp*_se*_emp*_adj*.dta

- 3a_micro_simulations_baseline.do

  This file builds a calculator to simulate crude horizontal equity and vertical equity measures across different rates of PIT and VAT. Then, simulates a budget neutral progressive reform that raises 1% of GDP from PIT rebated through VAT cut to measure changes in HE and VE trade-off.

  Inputs:

  $main/waste/${country_fullname}_VAT simulation data.dta
  $main/data/VAT_DATA.xlsx
  $main/proc/consumption_adjustment_factor.dta
  $main/data/TaxDevEITDv0.2.xlsx
  $main/data/isocodes_dictionary.dta
  $main/data/saving_rates.dta
  $survey/proc/${country_fullname}/${country_code}_${exp_file_name}.dta
  $survey/proc/${country_fullname}/${country_code}_labor_head.dta

$main/data/wb_stats.dta
$main/Notes/country_validation.dta

Outputs:

$main/waste/${country_code}_vat_heve_prep_exp*.dta
$main/proc/he_ve_meta_hh/${country_fullname}_linear rev_pit.dta
$main/proc/he_ve_meta_hh/${country_fullname}_linear rev_vat.dta
$main/proc/${country_code}_brackets_top_earners_exp*_se*_emp*_adj*.dta
$main/proc/he_ve_meta_hh/${country_fullname}_linear rev_pit_exp*_se*_emp*_adj*.dta
$main/proc/he_ve_meta_hh/${country_fullname}_linear rev_vat_exp*_se*_emp*_adj*.dta
$main/proc/he_ve_meta_hh/${country_fullname}_sim_reform_exp*_se*_emp*_adj*.dta
$main/graphs/${country_code}_ETR_exp*_se*_emp*_adj*.png
$main/graphs/${country_code}_income_exp*_se*_emp*_adj*.png
$main/proc/he_ve_meta_hh/crosscountry_sim_reform_exp*_se*_emp*_adj*.dta
$main/proc/he_ve_meta_hh/crosscountry_top_brackets_sim_reform_exp*_se*_emp*_adj*.dta

---

- `3b_micro_simulations_robustness.do`

  This file runs robustness checks to baseline micro simulations by systematically varying misreporting parameters for employees and self-employed workers across multiple scenarios, and testing alternative income definitions. It calls `3a_micro_simulations_baseline.do` repeatedly with different parameter combinations to assess sensitivity of horizontal and vertical equity measures. Specifically, it runs three misreporting scenarios (SE=50%/EMP=75%; SE=25%/EMP=75%; SE=25%/EMP=50%) and two alternative income definitions (total household expenditure; per-capita household expenditure).

  Inputs:

  Same inputs as `3a_micro_simulations_baseline.do`:

  $main/waste/${country_fullname}_VAT simulation data.dta
  $main/data/VAT_DATA.xlsx
  $main/proc/consumption_adjustment_factor.dta
  $main/data/TaxDevEITDv0.2.xlsx
  $main/data/isocodes_dictionary.dta
  $main/data/saving_rates.dta
  $survey/proc/${country_fullname}/${country_code}_${exp_file_name}.dta
  $survey/proc/${country_fullname}/${country_code}_labor_head.dta
  $main/data/wb_stats.dta
  $main/Notes/country_validation.dta

  Outputs:

  Same outputs as `3a_micro_simulations_baseline.do` but with varying parameter combinations (`se_compl`, `emp_compl`, `whichexp`):

  $main/waste/${country_code}_vat_heve_prep_exp*.dta
  $main/proc/he_ve_meta_hh/${country_fullname}_linear rev_pit.dta
  $main/proc/he_ve_meta_hh/${country_fullname}_linear rev_vat.dta
  $main/proc/${country_code}_brackets_top_earners_exp*_se*_emp*_adj*.dta
  $main/proc/he_ve_meta_hh/${country_fullname}_linear rev_pit_exp*_se*_emp*_adj*.dta

$main/proc/he_ve_meta_hh/${country_fullname}_linear rev_vat_exp*_se*_emp*_adj*.dta
$main/proc/he_ve_meta_hh/${country_fullname}_sim_reform_exp*_se*_emp*_adj*.dta
$main/graphs/${country_code}_ETR_exp*_se*_emp*_adj*.png
$main/graphs/${country_code}_income_exp*_se*_emp*_adj*.png
$main/proc/he_ve_meta_hh/crosscountry_sim_reform_exp*_se*_emp*_adj*.dta
$main/proc/he_ve_meta_hh/crosscountry_top_brackets_sim_reform_exp*_se*_emp*_adj*.dta

---

- 4a_tables.do

  This file produces descriptive and analytical tables for the paper using micro-simulation results. It generates four main tables: (1) descriptive statistics on countries in the sample including GDP, survey info, and tax parameters; (2) fitness of targeted moments comparing simulated and actual tax moments; (3) baseline micro-simulation results showing effective tax rates and horizontal/vertical equity measures; and (4) robustness checks under alternative misreporting assumptions. All tables are exported in LaTeX format.

  Inputs:

  $main/data/country_list_for_loop.xlsx
  $survey/proc/${country_fullname}/${country_code}_${exp_file_name}.dta
  $main/data/wb_stats.dta
  $main/data/VAT_DATA.xlsx
  $main/data/isocodes_dictionary.dta
  $main/data/TaxDevEITDv0.2.xlsx
  $main/data/ppp conversion factor.dta
  $main/proc/he_ve_meta_hh/crosscountry_sim_reform_exp*_se*_emp*_adj*.dta
  $main/proc/consumption_adjustment_factor.dta

  Outputs:

  $main/tables/cross_country_sample.tex
  $main/tables/targeted_moments_exp*_se*_emp*_adj*.tex
  $main/tables/he_ve_estimates_exp*_se*_emp*_adj*.tex
  $main/tables/he_ve_robustness_exp*_se*_emp*_adj*.tex

- 4b_figures.do

  This file produces all figures for the paper and appendix using micro-simulation results and cross-country data. It generates figures showing: (1) cross-country patterns for self-employment (share in top income decile and above PIT threshold); (2) consumption and VAT burden for self-employed and employees (informal consumption and effective VAT in top decile); (3) tax policy trade-offs of horizontal and vertical equity (scatter plots and reform arrows); (4) VAT and PIT collection across countries and income levels; (5) features of personal income tax (exemption thresholds and top marginal rates); (6) effective VAT rates and composition of imputed income; and (7) robustness figures for alternative income definitions. All figures are exported in PDF format.

  Inputs:

  $main/proc/final_employment_meta_exp*_se*_emp*_adj*.dta

$main/proc/final_metadata_exp*_se*_emp*_adj*.dta
$main/proc/final_consumption_meta_exp*_se*_emp*_adj*.dta
$main/proc/he_ve_meta_hh/crosscountry_sim_reform_exp*_se*_emp*_adj*.dta
$main/data/world-imf2024-v1.dta
$main/data/OGHIST_2025_10_07.xlsx
$main/data/wb_stats.dta

Outputs:

$main/graphs/share_se_p90_cross_count_exp*_se*_emp*_adj*.pdf
$main/graphs/share_se_pitexmp_cross_count_exp*_se*_emp*_adj*.pdf
$main/graphs/shares_inf_s_se_ci_exp*_se*_emp*_adj*.pdf
$main/graphs/eff_VAT_topdecile_exp*_se*_emp*_adj*.pdf
$main/graphs/he_ve_scatter_exp*_se*_emp*_adj*_v2.pdf
$main/graphs/he_ve_arrows_exp*_se*_emp*_adj*_v2.pdf
$main/graphs/cross_country_vat_pit_revenue.pdf
$main/graphs/cross_country_vat_pit_gdp_ratio.pdf
$main/graphs/topmtr_cross_count_exp*_se*_emp*_adj*.pdf
$main/graphs/exp_thres_cross_count_exp*_se*_emp*_adj*.pdf
$main/graphs/etr_above_1_cross_count_exp*_se*_emp*_adj*.pdf
$main/graphs/etr_above_5_cross_count_exp*_se*_emp*_adj*.pdf
$main/graphs/composition_imputed_income_exp*_se*_emp*_adj*.pdf
$main/graphs/vat_rates_exp*_se*_emp*_adj*.pdf
$main/graphs/he_ve_scatter_income_robustness.pdf
$main/graphs/he_ve_arrows_income_robustness.pdf

7.2 Codes for Experimental Survey Analysis

7.2.1 Online Survey

- **Do-file folder.** This folder contains all Stata `.do` scripts used to reproduce the online survey analysis. Scripts are numbered sequentially to indicate their execution order:
  - `1_master.do` – Main orchestrator. Sets global paths and macros, loads required packages, and runs all subsequent scripts in sequence.
  - `2_cleaning.do` – Imports and cleans the raw data; standardizes variable types and labels, handles missing values and duplicates, and saves cleaned files to `Data/Intermediate/`.
  - `3_regressions.do` – Runs all regressions (using `reghdfe`), equality tests, and control-mean calculations; exports publication-ready tables and figures to `Output/`.
  - `4_figures.do` – Produces heterogeneity figures (E4 and E5).
  - `5_tables.do` – Produces balance (E1) and demographic table (E2).
- **Output folder.** Contains all automatically generated outputs produced by the replication scripts. No files in this folder are manually edited. Outputs include regression tables (`.tex`), figures, summary statistics, and log files. Key outputs include:
  - **Table**
    - `Table E1: Balance checks by treatment arm` (country fixed effects included in all tests). This table is assembled from the following component files:
      - `balance_tp_all_arms.tex`
      - `balance_tp_all_arms_colombia.tex`
      - `balance_tp_all_arms_india.tex`
      - `balance_tp_all_arms_nigeria.tex`
      - `balance_tp_all_arms_philippines.tex`
      - `balance_tp_all_arms_indonesia.tex`
    - `Table D1: Online Survey: Randomization balance: E3_Table1withdem.tex`
    - `Table D2: Respondent characteristics by employment status: D2_online.tex`
    - `Table D3: Main regression results with q-values: D3_Table1withdem.tex`
  - **IMG**
    - `Figure D4`
    - `Figure D5`

7.2.2 In-Person Survey

- **Do-file folder.** This folder contains all Stata `.do` and R scripts used to reproduce the in-person survey analysis (including shared components with the online survey). Scripts are numbered sequentially to indicate their execution order:
  - `1_master.do` – Main orchestrator. Sets global paths and macros, loads required packages, and runs all subsequent scripts in sequence.
  - `2_cleaning.do` – Cleans and prepares the raw in-person and online survey data, merges stratification variables, harmonizes variable labels, and saves cleaned files for analysis.

- `3_regressions_overleaf.do` – Runs the main regressions and exports publication-ready LaTeX tables for Overleaf.
- `3_regressions_overleaf_nocontrols.do` – Runs the main regressions without control variables and exports formatted tables for Overleaf.
- `4_final_descriptives.do` – Produces descriptive statistics, summary tables, and graphs used in the paper.
- `5_multiple_hypothesis_updated.do` – Applies multiple-hypothesis testing adjustments (for example, sharpened q-values) to key outcome families.
- `6_causal forest figure.rtf` – Conducts causal forest exercise to analyze mediators of treatment effects.

- **Output folder.** Contains all automatically generated outputs produced by the replication scripts. No files in this folder are manually edited. Outputs include regression tables (`.tex`), figures, descriptive statistics, and graph files. Key outputs include:
  - **Table**
    - `Table 1: Treatment Effects on Perceptions of Unfairness (In-Person Survey)`
    - `Table 2: Treatment Effects on Perceptions of Horizontal Equity (In-Person Survey)`
    - `Table 3: Treatment Effects on Preferences for Sales Tax (In-Person Survey)`
    - `Table 4: Treatment Effects on Tax Morale (In-Person Survey)`
    - `Table 5: Heterogeneous Treatment Effects by Beliefs about Government Capacity`
    - `Table B1: Demographic Characteristics by Employment Status`
    - `Table B2: In-person Survey: Randomization balance`
    - `Table B3: Treatment Effects on Perceptions of Unfairness (Excluding Controls)`
    - `Table B4: Treatment Effects on Perceptions of Horizontal Equity(Excluding Controls)`
    - `Table B5: Treatment Effects on Preferences for Sales Tax(Excluding Controls)`
    - `Table B6: Treatment Effects on Tax Morale (Excluding Controls)`
    - `Table B7: Treatment Effects on Perceptions of Unfairness (Multiple Hypothesis Testing)`
    - `Table B8: Treatment Effects on Perceptions of Horizontal Equity (Multiple Hypothesis Testing)`
    - `Table B9: Treatment Effects on Preferences for Sales Tax (Multiple Hypothesis Testing)`
    - `Table B10: Treatment Effects on Tax Morale (Multiple Hypothesis Testing)`
  - **GPH**
    - `PanelA_combined.gph`: Figure 5, Panel A
    - `PanelB_combined.gph`: Figure 5, Panel B
    - `PanelC_combined.gph`: Figure 5, Panel C
  - **IMG**
    - `PanelA_combined.png`: Figure 5, Panel A

    * `PanelB_combined.png`: Figure 5, Panel B
    * `PanelC_combined.png`: Figure 5, Panel C
- **PDF**
    * `variable_importance_two_panel.pdf`: Figure B.7

7.3 Programs

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

- `blindschemes`
  Daniel Bischof. "BLINDSCHEMES: Stata module providing color-blind-friendly graph schemes."
- `carryforward`
  David Kantor, 2004. "CARRYFORWARD: Stata module to carry forward previous observations" Statistical Software Components S444902, Boston College Department of Economics, revised 12 Feb 2016.
- `cibar`
  Alexander Staudt, 2014. "CIBAR: Stata module to plot bar graphs and confidence intervals over groups" Statistical Software Components S457805, Boston College Department of Economics, revised 24 Jun 2019.
- `ciplot`
  Nicholas J. Cox, 2003. "CIPLOT: Stata module for plots of confidence intervals" Statistical Software Components S431202, Boston College Department of Economics, revised 15 Aug 2011.
- `coefplot`
  Ben Jann, 2013. "COEFPLOT: Stata module to plot regression coefficients and other results" Statistical Software Components S457686, Boston College Department of Economics, revised 22 Aug 2025.
- `colrspace`
  Ben Jann. "COLRSPACE: Stata module providing a Mata-based color management system for Stata graphics."
- `distinct`
  Gary Longton and Nicholas J. Cox. "DISTINCT: Stata module to display distinct values of variables" Statistical Software Components, Boston College Department of Economics.
- `egenmore`
  Nicholas J. Cox, 2000. "EGENMORE: Stata modules to extend the generate function" Statistical Software Components S386401, Boston College Department of Economics, revised 24 Jan 2019.
- `estout` (includes `esttab`, `eststo`, `estadd`)
  Ben Jann. "ESTOUT: Stata module to make regression tables from stored estimates" Statistical Software Components, Boston College Department of Economics.
- `fre`
  Ben Jann, 2007. "FRE: Stata module to display one-way frequency table" Statistical Software Components S456835, Boston College Department of Economics, revised 31 May 2021.
- `fs`
  Nicholas J. Cox. "FS: Stata module to show names of files in compact form" Statistical Software Components, Boston College Department of Economics.

- `ftools`
  Sergio Correia, 2016. "FTOOLS: Stata module to provide alternatives to common Stata commands optimized for large datasets" Statistical Software Components S458213, Boston College Department of Economics, revised 11 Jan 2026.

- `grc1leg`
  Vince Wiggins. "GRC1LEG: Stata module to combine multiple graphs with a single common legend."

- `grstyle`
  Ben Jann. "GRSTYLE: Stata module to customize the overall look of graphs" Statistical Software Components, Boston College Department of Economics.

- `gtools`
  Mauricio Caceres Bravo. "GTOOLS: Stata module providing a fast implementation of common group commands."

- `iebaltab`
  DIME Analytics, 2016. "IEBALTAB: Stata module to produce balance tables for impact evaluations" Statistical Software Components S458137, Boston College Department of Economics.

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

- `isocodes`
  Robert Picard. "ISOCODES: Stata module to clean ISO country codes" Statistical Software Components, Boston College Department of Economics.

- `labmask`
  Nicholas J. Cox. "LABMASK: Stata module for assignment of values as value labels" Statistical Software Components, Boston College Department of Economics.

- `labutil`
  Nicholas J. Cox. "LABUTIL: Stata modules for managing value and variable labels" Statistical Software Components, Boston College Department of Economics.

- `labvars`
  Nicholas J. Cox. "LABVARS: Stata module to create variable labels."

- `latexify`
  Sean Higgins. User-written package to convert matrix results to LaTeX tables row by row. Available at: https://github.com/skhiggins/StataTools

- `missings`
  Nicholas J. Cox, 2015. "MISSINGS: Stata module to manage missing values" Statistical Software Components S458071, Boston College Department of Economics, revised 11 May 2017.

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

- `mrgraph`
  Ben Jann. "MRGRAPH: utility to produce graphs of multiple-response variables" (distributed with the `mrtab` package).

- `mrtab`
  Ben Jann and Hilde Schaeper. "MRTAB: Stata module to compute one- and two-way tables of multiple responses" Statistical Software Components S437201, Boston College Department of Economics.

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

- `palettes`
  Ben Jann. "PALETTES: Stata module providing color, symbol, and line pattern palettes for Stata graphics."

- `reghdfe`
  Sergio Correia, 2017. "REGHDFE: Stata module for linear and instrumental-variable/GMM regression absorbing multiple levels of fixed effects" Statistical Software Components S457874, Boston College Department of Economics.

- `regsave`
  Julian Reif. "REGSAVE: Stata module to save regression results to a Stata-formatted dataset" Statistical Software Components, Boston College Department of Economics.

- `rscript`
  Reif, J. (2023).“Stata module to call an R script from Stata”. Available at: https://github.com/reifjulian/rscript

- `rwolf2`
  David A. Roodman. "RWOLF2: Stata module to compute Romano-Wolf stepdown p-values for multiple hypothesis testing" Statistical Software Components, Boston College Department of Economics.

- `schemepack`
  Asjad Naqvi, 2021. "SCHEMEPACK: Stata module providing ready-to-use graph schemes" Statistical Software Components S459012, Boston College Department of Economics, revised 11 Nov 2022.

- `slideplot`
  Nicholas J. Cox, 2003. "SLIDEPLOT: Stata module for sliding bar plot for frequencies or percents" Statistical Software Components S432801, Boston College Department of Economics.

- `stripplot`
  Nicholas J. Cox. "STRIPPLOT: Stata module for strip plots."

- `tabout`
  Ian Watson, 2004. "TABOUT: Stata module to export publication quality cross-tabulations"

Statistical Software Components S447101, Boston College Department of Economics, revised 16 Mar 2019.

- `tabstatmat`
  Nicholas J. Cox. "TABSTATMAT: Stata module to save matrices saved by tabstat to single named matrix" Statistical Software Components, Boston College Department of Economics.

- `unique`
  Tony Brady and Nicholas J. Cox. "UNIQUE: Stata module to report number of unique values in variable(s)" Statistical Software Components, Boston College Department of Economics.

- `valuesof`
  Nicholas J. Cox. "VALUESOF: Stata module to return values of a variable" Statistical Software Components, Boston College Department of Economics.

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

- `xfill`
  Sealed Envelope Ltd. "XFILL: Stata module for forward and backward filling of missing values." Available at: https://www.sealedenvelope.com/stata/xfill/

A Appendix: Secondary Codes

This section provides a detailed description of each do-file that cleans the labor and expenditure modules of the household survey data for each of the 28 countries in the sample, producing harmonized, clean data files that are subsequently used for micro-simulations. These codes are invoked in the do file 1b_cleaning_hh_surveys.do

A.1 Codes Cleaning Labor Modules

- BI_2014_ECVM_labor.do
  Take raw household survey data from Burundi and harmonize labor module to produce clean final working data file.

  Inputs:
  $survey/raw/Burundi2014/individu.dta

  Outputs:
  $survey/proc/Burundi2014/BI_labor_head.dta

- BF_2009_EICVM_labor.do
  Take raw household survey data from Burkina Faso and harmonize labor module to produce clean final working data file.

  Inputs:
  $survey/raw/BurkinaFaso2009/poverty_final_version_13.dta

  Outputs:
  $survey/proc/BurkinaFaso2009/BF_labor_head.dta

- BJ_2015_EICVN_labor.do
  Take raw household survey data from Benin and harmonize labor module to produce clean final working data file.

  Inputs:
  $survey/raw/Benin2015/individu.dta

  Outputs:
  $survey/proc/Benin2015/BJ_labor_head.dta

- BR_2009_POF_labor.do
  Take raw household survey data from Brazil and harmonize labor module to produce clean final working data file.

  Inputs:
  $survey/raw/Brazil2009/rend_dedu.dta
  $survey/raw/Brazil2009/head_covariates.dta

Outputs:

$survey/proc/Brazil2009/BR_labor_head.dta

- CD_2005_E123_labor.do

  Take raw household survey data from Congo DRC and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Congo_DRC2005/Phase1_IND.dta

  Outputs:

  $survey/proc/Congo_DRC2005/CD_labor_head.dta

- CG_2005_ECOM_labor.do

  Take raw household survey data from Congo Republic and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Congo_Rep2005/ecom2005ind2.dta

  $survey/raw/Congo_Rep2005/ecom2005men.dta

  Outputs:

  $survey/proc/Congo_Rep2005/CG_labor_head.dta

- CL_2017_EPF_labor.do

  Take raw household survey data from Chile and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Chile2017/BASE_PERSONAS_VIII_EPF.dta

  Outputs:

  $survey/proc/Chile2017/CL_labor_head.dta

- CM_2014_ECAM_labor.do

  Take raw household survey data from Cameroon and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Cameroon2014/ECAM4IND.dta

  Outputs:

  $survey/proc/Cameroon2014/CM_labor_head.dta

- CO_2007_ENIG_labor.do

  Take raw household survey data from Colombia and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Colombia2007/Ig_ml_ocupado.dta
  $survey/raw/Colombia2007/Ig_ml_persona.dta
  $survey/raw/Colombia2007/Ig_ml_gas_dia.dta
  $survey/raw/Colombia2007/Ig_ml_gas_sem.dta

  Outputs:

  $survey/proc/Colombia2007/CO_labor_head.dta

- CR_2014_ENIGH_labor.do

  Take raw household survey data from Costa Rica and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Costa_Rica2014/ENIGH_2013_BASE_PERSONAS.dta
  $survey/raw/Costa_Rica2014/ENIGH_2013_BASE_HOGARES.dta

  Outputs:

  $survey/proc/Costa_Rica2014/CR_labor_head.dta

- DO_2007_ENIGH_labor.do

  Take raw household survey data from Dominican Republic and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Dominican_Rep2007/02 cuestionario 1_Poblacion.dta
  $survey/raw/Dominican_Rep2007/01 Datos_Hogar.dta

  Outputs:

  $survey/proc/Dominican_Rep2007/DO_labor_head.dta

- EC_2012_ENIGHUR_labor.do

  Take raw household survey data from Ecuador and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Ecuador2012/04_enighur11_personas.dta
  $survey/raw/Ecuador2012/01_enighur11_identificacion.dta

  Outputs:

  $survey/proc/Ecuador2012/EC_labor_head.dta

- ID_2019_SUSENAS_labor.do

  Take raw household survey data from Indonesia and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Indonesia2019/kor19ind_1_diseminasi.dta

  $survey/raw/Indonesia2019/kor19rt_revisi1_diseminasi.dta

  Outputs:

  $survey/proc/Indonesia2019/ID_labor_head.dta

- IN_2012_NSS_labor.do

  Take raw household survey data from India and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/India2012/Household characteristics - Block 3 - Level2 - type2 - 68.dta

  $survey/raw/India2012/Household characteristics - Block 3 - level3 - type2 - 68.dta

  $survey/raw/India2012/Demographic and other particulars of household members - Block 4 - Level 4 - Type 2 - 68.dta

  Outputs:

  $survey/proc/India2012/IN_labor_head.dta

- MA_2001_ENCDM_labor.do

  Take raw household survey data from Morocco and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Morocco2001/fichier_individu_encdm2000_01.dta

  Outputs:

  $survey/proc/Morocco2001/MA_labor_head.dta

- ME_2009_HBS_labor.do

  Take raw household survey data from Montenegro and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Montenegro2009/indiv_labelled.dta

  $survey/raw/Montenegro2009/general_info.dta

  Outputs:

  $survey/proc/Montenegro2009/ME_labor_head.dta

- MX_2014_ENIGH_labor.do

  Take raw household survey data from Mexico and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Mexico2014/ncv_poblacion_2014_concil_2010_stata.dta
  $survey/raw/Mexico2014/tra_trabajos_2014_concil_2010.dta
  $survey/raw/Mexico2014/ncv_concentrado_2014_concil_2010_stata.dta
  $survey/raw/Mexico2014/ncv_vivi_2014_concil_2010_stata.dta

  Outputs:

  $survey/proc/Mexico2014/MX_labor_head.dta

- MZ_2009_IOF_labor.do

  Take raw household survey data from Mozambique and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Mozambique2009/agregado_v3.dta
  $survey/raw/Mozambique2009/iof_ponderadores3.dta
  $survey/raw/Mozambique2009/controle_v3.dta

  Outputs:

  $survey/proc/Mozambique2009/MZ_labor_head.dta

- NE_2007_ENBC_labor.do

  Take raw household survey data from Niger and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Niger2007/menageidentifiant.dta
  $survey/raw/Niger2007/r05a1.dta
  $survey/raw/Niger2007/r05c.dta
  $survey/raw/Niger2007/r01.dta
  $survey/raw/Niger2007/r02.dta

  Outputs:

  $survey/proc/Niger2007/NE_labor_head.dta

- PE_2017_ENAHO_labor.do

  Take raw household survey data from Peru and harmonize labor module to produce clean final working data file.

Inputs:

$survey/raw/Peru2017/enaho01-2017-500.dta
$survey/raw/Peru2017/enaho01-2017-200.dta
$survey/raw/Peru2017/enaho01a-2017-300.dta
$survey/raw/Peru2017/enaho01-2017-100.dta

Outputs:

$survey/proc/Peru2017/PE_labor_head.dta

- PH_2018_FIES_labor.do

  Take raw household survey data from Philippines and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Philippines2018/2018_FIES_vol2.csv
  $survey/raw/Philippines2018/FIES PUF 2018 Vol.1.csv

  Outputs:

  $survey/proc/Philippines2018/PH_labor_head.dta

- PK_2019_PSLM_labor.do

  Take raw household survey data from Pakistan and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Pakistan2019/plist.dta
  $survey/raw/Pakistan2019/sec_5a.dta
  $survey/raw/Pakistan2019/sec_2ab.dta
  $survey/raw/Pakistan2019/sec_1b (2).dta

  Outputs:

  $survey/proc/Pakistan2019/PK_labor_head.dta

- PY_2011_EIGyCV_labor.do

  Take raw household survey data from Paraguay and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Paraguay2011/Datos de la poblacion reg02t.dta
  $survey/raw/Paraguay2011/Datos de la vivienda y del Hogar reg01.dta

  Outputs:

  $survey/proc/Paraguay2011/PY_labor_head.dta

- RW_2014_EICV_labor.do

  Take raw household survey data from Rwanda and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Rwanda2014/cs_s6b_employement_6c_salaried_s6d_business.dta

  $survey/raw/Rwanda2014/cs_s1_s2_s3_s4_s6a_s6e_s6f_person.dta

  Outputs:

  $survey/proc/Rwanda2014/RW_labor_head.dta

- SN_2008_EDMC_labor.do

  Take raw household survey data from Senegal and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Senegal_Dakar2008/CQ06Aan.dta

  $survey/raw/Senegal_Dakar2008/CQ03Aan.dta

  $survey/raw/Senegal_Dakar2008/CQ02Aan.dta

  Outputs:

  $survey/proc/Senegal_Dakar2008/SN_labor_head.dta

- TN_2010_EBC_labor.do

  Take raw household survey data from Tunisia and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Tunisia2010/menage.dta

  $survey/raw/Tunisia2010/bdg2010T2.dta

  Outputs:

  $survey/proc/Tunisia2010/TN_labor_head.dta

- TZ_2012_HBS_labor.do

  Take raw household survey data from Tanzania and harmonize labor module to produce clean final working data file.

  Inputs:

  $survey/raw/Tanzania2012/SECT12.dta

  $survey/raw/Tanzania2012/SECT1.dta

  Outputs:

  $survey/proc/Tanzania2012/TZ_labor_head.dta

A.2 Codes Cleaning Expenditure Modules

- BF_2009_EICVM.do

  Take raw household survey data from Burkina Faso and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/BurkinaFaso2009/EICVM_fichier_menage_produit.dta
  $survey/raw/BurkinaFaso2009/CQ02.dta
  $survey/raw/BurkinaFaso2009/CQ04.dta
  $survey/raw/BurkinaFaso2009/CQ11.dta
  $survey/raw/BurkinaFaso2009/CQ05.dta
  $survey/raw/BurkinaFaso2009/CQ06.dta
  $survey/raw/BurkinaFaso2009/CQ07.dta
  $survey/raw/BurkinaFaso2009/CQ08.dta
  $survey/raw/BurkinaFaso2009/CQ09.dta
  $survey/raw/BurkinaFaso2009/CQ10.dta
  $survey/raw/BurkinaFaso2009/CQ11.dta
  $survey/raw/BurkinaFaso2009/CQ12.dta
  $survey/raw/BurkinaFaso2009/CQ13.dta
  $survey/raw/BurkinaFaso2009/CQ11.dta
  $survey/raw/BurkinaFaso2009/CQ14.dta
  $survey/raw/BurkinaFaso2009/CQ15.dta
  $survey/raw/BurkinaFaso2009/CQ16.dta
  $survey/raw/BurkinaFaso2009/CQ17.dta
  $survey/raw/BurkinaFaso2009/CQ18.dta
  $survey/raw/BurkinaFaso2009/CQ19.dta
  $survey/raw/BurkinaFaso2009/CQ20.dta
  $survey/raw/BurkinaFaso2009/CQ21.dta
  $survey/raw/BurkinaFaso2009/CQ22.dta
  $survey/raw/BurkinaFaso2009/CQ23.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/BF_household_cov_original.dta
  $survey/proc/BF_household_cov.dta
  $survey/raw/BurkinaFaso2009/expenses/adj_CQ02.dta
  $survey/raw/BurkinaFaso2009/expenses/adj_CQ04.dta
  $survey/raw/BurkinaFaso2009/expenses/adj_CQ05.dta
  $survey/raw/BurkinaFaso2009/expenses/adj_CQ06.dta
  $survey/raw/BurkinaFaso2009/expenses/adj_CQ07.dta
  $survey/raw/BurkinaFaso2009/expenses/adj_CQ08.dta
  $survey/raw/BurkinaFaso2009/expenses/adj_CQ09.dta
  $survey/raw/BurkinaFaso2009/expenses/adj_CQ10.dta

$survey/raw/BurkinaFaso2009/expenses/adj_CQ11.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ12.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ13.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ14.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ15.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ16.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ17.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ18.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ19.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ20.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ21.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ22.dta
$survey/raw/BurkinaFaso2009/expenses/adj_CQ23.dta
$survey/waste/BF_all_lines_raw.dta
$survey/proc/BurkinaFaso2009/BurkinaFaso2009_TOR_stats_for_crosswalk.xls
$survey/proc/BurkinaFaso2009/BF_exp_full_db.dta
$survey/proc/BurkinaFaso2009/BF_temp_exp_TOR_item_COICOP_4dig.dta

---

- BI_2014_ECVM.do

  Take raw household survey data from Burundi and harmonize expenditure module to produce clean final working data file.

  Inputs:
  $survey/raw/Burundi2014/poverty_final_version_13.dta
  $survey/raw/Burundi2014/CQ03.dta
  $survey/raw/Burundi2014/CQ04.dta
  $survey/raw/Burundi2014/CQ05.dta
  $survey/raw/Burundi2014/CQ06.dta
  $survey/raw/Burundi2014/CQ07.dta
  $survey/raw/Burundi2014/CQ08.dta
  $survey/raw/Burundi2014/CQ09.dta
  $survey/raw/Burundi2014/CQ10.dta
  $survey/raw/Burundi2014/CQ11.dta
  $survey/raw/Burundi2014/CQ12.dta
  $survey/raw/Burundi2014/CQ13.dta
  $survey/raw/Burundi2014/CQ14.dta
  $survey/raw/Burundi2014/CQ15.dta
  $survey/raw/Burundi2014/CQ16.dta
  $survey/raw/Burundi2014/CQ17.dta
  $survey/raw/Burundi2014/CQ18.dta
  $survey/raw/Burundi2014/CQ19.dta
  $survey/raw/Burundi2014/CQ20.dta
  $survey/raw/Burundi2014/CQ21.dta
  $survey/raw/Burundi2014/CQ22.dta

$survey/raw/Burundi2014/CQ23.dta
$survey/raw/additional/COICOP_label_2dig.dta
$survey/raw/additional/COICOP_label_3dig.dta
$survey/raw/additional/COICOP_label_4dig.dta

Outputs:

$survey/waste/BI_household_cov_original.dta
$survey/proc/Burundi2014/BI_household_cov.dta
$survey/raw/Burundi2014/adj_CQ03.dta
$survey/raw/Burundi2014/adj_CQ04.dta
$survey/raw/Burundi2014/adj_CQ05.dta
$survey/raw/Burundi2014/adj_CQ06.dta
$survey/raw/Burundi2014/adj_CQ07.dta
$survey/raw/Burundi2014/adj_CQ08.dta
$survey/raw/Burundi2014/adj_CQ09.dta
$survey/raw/Burundi2014/adj_CQ10.dta
$survey/raw/Burundi2014/adj_CQ11.dta
$survey/raw/Burundi2014/adj_CQ12.dta
$survey/raw/Burundi2014/adj_CQ13.dta
$survey/raw/Burundi2014/adj_CQ14.dta
$survey/raw/Burundi2014/adj_CQ15.dta
$survey/raw/Burundi2014/adj_CQ16.dta
$survey/raw/Burundi2014/adj_CQ17.dta
$survey/raw/Burundi2014/adj_CQ18.dta
$survey/raw/Burundi2014/adj_CQ19.dta
$survey/raw/Burundi2014/adj_CQ20.dta
$survey/raw/Burundi2014/adj_CQ21.dta
$survey/raw/Burundi2014/adj_CQ22.dta
$survey/raw/Burundi2014/adj_CQ23.dta
$survey/waste/BI_all_lines_raw.dta
$survey/proc/Burundi2014/Burundi2014_TOR_stats_for_crosswalk.xls
$survey/proc/Burundi2014/BI_exp_full_db.dta
$survey/proc/Burundi2014/BI_temp_exp_TOR_item_COICOP_4dig.dta

---

- BJ_2015_EICVN.do

  Take raw household survey data from Benin and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Benin2015/individu.dta
  $survey/raw/Benin2015/RM11.dta
  $survey/raw/Benin2015/RMAL.dta
  $survey/raw/Benin2015/RM12.dta
  $survey/raw/Benin2015/RM12a.dta
  $survey/raw/Benin2015/RM13.dta

$survey/raw/Benin2015/RM13a.dta
$survey/raw/Benin2015/RM14.dta
$survey/raw/Benin2015/RM14a.dta
$survey/raw/Benin2015/RM15.dta
$survey/raw/Benin2015/RM15a.dta
$survey/raw/Benin2015/RM16.dta
$survey/raw/Benin2015/RM16a.dta
$survey/raw/Benin2015/RM17.dta
$survey/raw/Benin2015/RM17a.dta
$survey/raw/Benin2015/RM18.dta
$survey/raw/Benin2015/RM18a.dta
$survey/raw/Benin2015/RM19.dta
$survey/raw/Benin2015/RM19a.dta
$survey/raw/Benin2015/RM20.dta
$survey/raw/Benin2015/RM20a.dta
$survey/raw/Benin2015/RM21.dta
$survey/raw/Benin2015/RM21a.dta
$survey/raw/additional/COICOP_label_2dig.dta
$survey/raw/additional/COICOP_label_3dig.dta
$survey/raw/additional/COICOP_label_4dig.dta

Outputs:

$survey/raw/Benin2015/adj_RM11.dta
$survey/raw/Benin2015/adj_RMAL.dta
$survey/raw/Benin2015/adj_RM12.dta
$survey/raw/Benin2015/adj_RM12a.dta
$survey/raw/Benin2015/adj_RM13.dta
$survey/raw/Benin2015/adj_RM13a.dta
$survey/raw/Benin2015/adj_RM14.dta
$survey/raw/Benin2015/adj_RM14a.dta
$survey/raw/Benin2015/adj_RM15.dta
$survey/raw/Benin2015/adj_RM15a.dta
$survey/raw/Benin2015/adj_RM16.dta
$survey/raw/Benin2015/adj_RM16a.dta
$survey/raw/Benin2015/adj_RM17.dta
$survey/raw/Benin2015/adj_RM17a.dta
$survey/raw/Benin2015/adj_RM18.dta
$survey/raw/Benin2015/adj_RM18a.dta
$survey/raw/Benin2015/adj_RM19.dta
$survey/raw/Benin2015/adj_RM19a.dta
$survey/raw/Benin2015/adj_RM20.dta
$survey/raw/Benin2015/adj_RM20a.dta
$survey/raw/Benin2015/adj_RM21.dta
$survey/waste/BJ_all_lines_raw.dta
$survey/proc/Benin2015/Benin2015_TOR_stats_for_crosswalk.xls

$survey/proc/Benin2015/BJ_exp_full_db.dta
$survey/proc/Benin2015/BJ_temp_exp_TOR_item_COICOP_4dig.dta

- BR_2009_POF.do

  Take raw household survey data from Brazil and harmonize expenditure module to produce clean final working data file.

  Inputs:
  $survey/raw/Brazil2009/t_morador_s.dta
  $survey/raw/Brazil2009/t_domicilio_s.dta
  $survey/raw/Brazil2009/BR_POF_crosswalk_TOR_all.csv
  $survey/raw/Brazil2009/BR_POF_crosswalk_product_code.csv
  $survey/raw/Brazil2009/t_despesa_12meses_s.dta
  $survey/raw/Brazil2009/t_despesa_90dias_s.dta
  $survey/raw/Brazil2009/t_despesa_veiculo_s.dta
  $survey/raw/Brazil2009/t_caderneta_despesa_s.dta
  $survey/raw/Brazil2009/t_despesa_individual_s.dta
  $survey/raw/Brazil2009/t_outras_despesas_s.dta
  $survey/raw/Brazil2009/t_servico_doms_s.dta
  $survey/raw/Brazil2009/t_aluguel_estimado_s.dta
  $survey/raw/Brazil2009/BR_POF_crosswalk_COICOP_all_category_codes.dta
  $survey/raw/Brazil2009/BR_POF_crosswalk_TOR_3dig.xlsx
  $survey/waste/all_lines_no_TOR.dta
  $survey/raw/additional/COICOP_label_2dig.dta

  Outputs:
  $survey/waste/BR_household_cov_original.dta
  $survey/proc/Brazil2009/BR_household_cov.dta
  $survey/raw/Brazil2009/all_retailers_codes.dta
  $survey/raw/Brazil2009/product_codes.dta
  $survey/waste/BR_all_lines_raw.dta
  $survey/proc/Brazil2009/Brazil2009_TOR_stats_for_crosswalk.xls
  $survey/proc/Brazil2009/BR_exp_full_db.dta
  $survey/proc/Brazil2009/BR_temp_exp_TOR_item_product_code_4dig.dta

---

- CD_2005_E123.do

  Take raw household survey data from Congo DRC and harmonize expenditure module to produce clean final working data file.

  Inputs:
  $survey/raw/Congo_DRC2005/Phase1_IND.dta
  $survey/raw/Congo_DRC2005/Phase3_final_depenses.dta
  $survey/raw/additional/COICOP_label_2dig.dta

$survey/raw/additional/COICOP_label_3dig.dta
$survey/raw/additional/COICOP_label_4dig.dta

Outputs:

$survey/waste/CD_household_cov_original.dta
$survey/proc/Congo_DRC2005/CD_household_cov.dta
$survey/waste/CD_all_lines_raw.dta
$survey/proc/Congo_DRC2005/Congo_DRC2005_TOR_stats_for_crosswalk.xls
$survey/proc/Congo_DRC2005/CD_exp_full_db.dta
$survey/proc/Congo_DRC2005/CD_temp_exp_TOR_item_COICOP_4dig.dta

---

- CG_2005_ECOM.do
  Take raw household survey data from Congo Rep. and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Congo_Rep2005/ecom2005men.dta
  $survey/raw/Congo_Rep2005/ecom2005pivot.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/CG_household_cov_original.dta
  $survey/proc/Congo_Rep2005/CG_household_cov.dta
  $survey/waste/CG_all_lines_raw.dta
  $survey/proc/Congo_Rep2005/Congo_Rep2005_TOR_stats_for_crosswalk.xls
  $survey/proc/Congo_Rep2005/CG_exp_full_db.dta
  $survey/proc/Congo_Rep2005/CG_temp_exp_TOR_item_COICOP_4dig.dta

---

- CL_2017_EPF.do
  Take raw household survey data from Chile and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Chile2017/BASE_PERSONAS_VIII_EPF.dta
  $survey/raw/Chile2017/BASE_ESTABLECIMIENTOS_VIII_EPF.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/CL_household_cov_original.dta
  $survey/proc/Chile2017/CL_household_cov.dta
  $survey/waste/CL_all_lines_raw.dta

$survey/proc/Chile2017/CL_TOR_stats_for_crosswalk.csv
$survey/proc/Chile2017/CL_exp_full_db.dta
$survey/proc/Chile2017/CL_temp_exp_TOR_item_COICOP_4dig.dta

- CM_2014_ECAM.do
  Take raw household survey data from Cameroon and harmonize expenditure module to produce clean final working data file.

  Inputs:
  $survey/raw/Cameroon2014/ECAM4IND.dta
  $survey/raw/Cameroon2014/ECAM4MEN.dta
  $survey/raw/Cameroon2014/ECAM4_Weight_12052015.dta
  $survey/raw/Cameroon2014/ECAM4LOG.dta
  $survey/raw/Cameroon2014/hh_income.dta
  $survey/raw/Cameroon2014/ECAM4PRODUIT.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:
  $survey/waste/CM_household_cov_original.dta
  $survey/proc/Cameroon2014/CM_household_cov.dta
  $survey/waste/CM_all_lines_raw_diary_no_selfprod.dta
  $survey/waste/CM_durables_with_imputed_TOR.dta
  $survey/waste/CM_all_lines_raw.dta
  $survey/proc/Cameroon2014/CM_TOR_stats_for_crosswalk.xls
  $survey/proc/Cameroon2014/CM_exp_full_db.dta
  $survey/proc/Cameroon2014/CM_temp_exp_TOR_item_COICOP_4dig.dta

---

- CO_2007_ENIG.do
  Take raw household survey data from Colombia and harmonize expenditure module to produce clean final working data file.

  Inputs:
  $survey/raw/Colombia2007/Ig_ml_persona.dta
  $survey/raw/Colombia2007/Ig_gsdu_gas_dia.dta
  $survey/raw/Colombia2007/Ig_gssr_gas_sem.dta
  $survey/raw/Colombia2007/Ig_ml_hogar.dta
  $survey/raw/Colombia2007/CO_dwellings_original.dta
  $survey/raw/Colombia2007/Ig_gsdp_gas_dia.dta
  $survey/raw/Colombia2007/Ig_gsmf_compra.dta
  $survey/raw/Colombia2007/Ig_gsmf_serv_pub.dta
  $survey/raw/Colombia2007/Ig_gsdu_gas_dia.dta

$survey/raw/additional/COICOP_label_2dig.dta
$survey/raw/additional/COICOP_label_3dig.dta
$survey/raw/additional/COICOP_label_4dig.dta

Outputs:

$survey/waste/CO_household_cov_original.dta
$survey/waste/CO_all_lines_raw.dta
$survey/waste/CO_household_cov_original.dta
$survey/proc/Colombia2007/CO_household_cov.dta
$survey/proc/Colombia2007/Colombia2007_TOR_stats_for_crosswalk.xls
$survey/proc/Colombia2007/CO_exp_full_db.dta
$survey/proc/Colombia2007/CO_temp_exp_TOR_item_COICOP_4dig.dta

---

- CR_2014_ENIGH.do

  Take raw household survey data from Costa Rica and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/proc/Costa_Rica2014/ENIGH_2013_BASE_HOGARES.dta
  $survey/proc/Costa_Rica2014/ENIGH_2013_BASE_GASTOS_EXTENDIDA_PUBLICA.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/CR_household_cov_original.dta
  $survey/waste/CR_all_lines_raw.dta
  $survey/proc/Costa_Rica2014/CR_household_cov.dta
  $survey/proc/Costa_Rica2014/Costa_Rica2014_TOR_stats_for_crosswalk.xls
  $survey/proc/Costa_Rica2014/CR_exp_full_db.dta
  $survey/proc/Costa_Rica2014/CR_temp_exp_TOR_item_COICOP_4dig.dta

---

- DO_2007_ENIGH.do

  Take raw household survey data from Dominican Republic and expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Dominican_Rep2007/02 cuestionario 1_Poblacion.dta
  $survey/raw/Dominican_Rep2007/01 Datos_Hogar.dta
  $survey/raw/Dominican_Rep2007/03 cuestionarios 1-2-3_Gasto.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/DO_household_cov_original.dta

$survey/waste/DO_all_lines_raw.dta
$survey/proc/Dominican_Rep2007/DO_household_cov.dta
$survey/proc/Dominican_Rep2007/Dominican_Rep2007_TOR_stats_for_crosswalk.xls
$survey/proc/Dominican_Rep2007/DO_exp_full_db.dta
$survey/proc/Dominican_Rep2007/DO_temp_exp_TOR_item_COICOP_4dig.dta

---

- EC_2012_ENIGHUR.do

  Take raw household survey data from Ecuador and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Ecuador2012/04_enighur11_personas.dta
  $survey/raw/Ecuador2012/01_enighur11_identificacion.dta
  $survey/raw/Ecuador2012/28_enighur11_gdiarios_seccion2.dta
  $survey/raw/Ecuador2012/29_enighur11_gdiarios_seccion3.dta
  $survey/raw/Ecuador2012/30_enighur11_gdiarios_seccion4.dta
  $survey/raw/Ecuador2012/35_enighur11_gmensuales.dta
  $survey/raw/Ecuador2012/36 enighur11_gtrimestrales.dta
  $survey/raw/Ecuador2012/37_enighur11_gsemestrales.dta
  $survey/raw/Ecuador2012/38_enighur11_ganuales.dta
  $survey/raw/Ecuador2012/31_enighur11_gmensuales_partea.dta
  $survey/raw/Ecuador2012/32_enighur11_gmensuales_parteb.dta
  $survey/raw/Ecuador2012/33_enighur11_gmensuales_partec.dta
  $survey/raw/Ecuador2012/34_enighur11_gmensuales_parted.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/EC_household_cov_original.dta
  $survey/waste/EC_all_lines_raw.dta
  $survey/proc/Ecuador2012/EC_household_cov.dta
  $survey/proc/Ecuador2012/Ecuador2012_TOR_stats_for_crosswalk.xls
  $survey/proc/Ecuador2012/EC_exp_full_db.dta
  $survey/proc/Ecuador2012/EC_temp_exp_TOR_item_COICOP_4dig.dta

---

- ID_2019_SUSENAS.do

  Take raw household survey data from Indonesia and harmonize module to produce clean final working data file.

  Inputs:

  $survey/raw/Indonesia2019/kor19rt_revisi1_diseminasi.dta
  $survey/raw/Indonesia2019/kor19ind_1_diseminasi.dta
  $survey/raw/Indonesia2019/blok41_gab_diseminasi_prov11_32rev.dta
  $survey/raw/Indonesia2019/blok41_gab_diseminasi_prov33_65rev.dta

$survey/raw/Indonesia2019/blok41_gab_diseminasi_prov71_94rev.dta
$survey/raw/Indonesia2019/blok42_revisi_diseminasi_prov11_36
$survey/raw/Indonesia2019/blok42_revisi_diseminasi_prov51_94

Outputs:

$survey/waste/ID_household_cov_original.dta
$survey/proc/Indonesia2019/ID_temp_exp.dta

- IN_2012_NSS.do

  Take raw household survey data from India and harmonize module to produce clean final working data file.

  Inputs:

  $survey/raw/India2012/Household characteristics - Block 3 - Level2 - type2 - 68.dta
  $survey/raw/India2012/Household characteristics - Block 3 - level3 - type2 - 68.dta
  $survey/raw/India2012/Block 4 - Level 4 - Type 2 - 68.dta
  $survey/raw/India2012/Block 5.1- 5.2- 6 - Level 5 - type 2 - 68.dta
  $survey/raw/India2012/Block 7 and 8 - Level 6 - Type 2 - 68.dta
  $survey/raw/India2012/Block 9 - Level 7 - type 2 - 68.dta
  $survey/raw/India2012/Block 10 - Level 8 - type 2 - 68.dta
  $survey/raw/India2012/Block 11 - Level 9 - Type 2 - 68.dta

  Outputs:

  $survey/waste/IN_household_cov_original.dta
  $survey/proc/India2012/IN_temp_exp.dta

- KM_2013_EDMC.do

  Take raw household survey data from Comoros and harmonize module to produce clean final working data file.

  Inputs:

  $survey/raw/Comoros2013/demo.dta
  $survey/raw/Comoros2013/apcq03 à 21.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/KM_household_cov_original.dta
  $survey/waste/KM_all_lines_raw.dta
  $survey/proc/Comoros2013/KM_household_cov.dta
  $survey/proc/Comoros2013/Comoros2013_TOR_stats_for_crosswalk.xls
  $survey/proc/Comoros2013/KM_exp_full_db.dta
  $survey/proc/Comoros2013/KM_temp_exp_TOR_item_COICOP_4dig.dta

---

- MA_2001_ENCDM.do

  Take raw household survey data from Morocco and harmonize module to produce clean final working data file.

  Inputs:

  $survey/raw/Morocco2001/ichier_individu_encdm2000_01.dta
  $survey/raw/Morocco2001/BM_with_hhid.dta
  $survey/raw/Morocco2001/expenses_and_totals.dta
  $survey/raw/Morocco2001/MA_ENCDM_crosswalk_COICOP.csv
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/MA_household_cov_original.dta
  $survey/waste/MA_all_lines_raw.dta
  $survey/proc/Morocco2001/MA_household_cov.dta
  $survey/proc/Morocco2001/Morocco2001_TOR_stats_for_crosswalk.xls
  $survey/proc/Morocco2001/MA_exp_full_db.dta
  $survey/proc/Morocco2001/MA_temp_exp_TOR_item_COICOP_4dig.dta
  $survey/proc/Morocco2001/Morocco_codes_1dig.dta
  $survey/proc/Morocco2001/Morocco_codes_2dig.dta
  $survey/proc/Morocco2001/Morocco_codes_3dig.dta
  $survey/waste/$country_code_all_lines_with_crosswalk.dta
  $survey/waste/codes_from_data_not_in_list.dta
  $survey/proc/Morocco2001/product_codes_levels_3_4.dta
  $survey/proc/Morocco2001/product_codes.dta

- ME_2009_HBS.do

  Take raw household survey data from Montenegro and harmonize module to produce clean final working data file.

  Inputs:

  $survey/raw/Montenegro2009/indiv_labelled.dta
  $survey/raw/Montenegro2009/general_info.dta
  $survey/raw/Montenegro2009/diary_labelled.dta
  $survey/raw/Montenegro2009/p2izdaci.dta
  $survey/raw/Montenegro2009/p2stan.dta
  $survey/raw/Montenegro2009/p3kupovina.dta
  $survey/raw/Montenegro2009/diary_labelled.dta
  $survey/raw/Montenegro2009/nonfood items.dta
  $survey/waste/ME_durables.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

$survey/waste/ME_household_cov_original.dta
$survey/waste/ME_all_lines_raw.dta
$survey/proc/Montenegro2009/ME_household_cov.dta
$survey/proc/Montenegro2009/Montenegro2009_TOR_stats_for_crosswalk.xls
$survey/proc/Montenegro2009/ME_exp_full_db.dta
$survey/proc/Montenegro2009/ME_temp_exp_TOR_item_COICOP_4dig.dta
$survey/waste/ME_durables.dta
$survey/waste/ME_durables_with_imputed_TOR.dta

- MX_2014_ENIGH.do
  Take raw household survey data from Mexico and harmonize module to produce clean final working data file.

  Inputs:
  $survey/raw/Mexico2014/ncv_concentrado_2014_concil_2010_stata.dta
  $survey/raw/Mexico2014/ncv_gastohogar_2014_concil_2010_stata.dta
  $survey/raw/Mexico2014/MX_ENIGH_crosswalk_COICOP.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:
  $survey/waste/MX_household_cov_original.dta
  $survey/waste/MX_all_lines_raw.dta
  $survey/proc/Mexico2014/MX_household_cov.dta
  $survey/proc/Mexico2014/Mexico2014_TOR_stats_for_crosswalk.xls
  $survey/proc/Mexico2014/MX_exp_full_db.dta
  $survey/proc/Mexico2014/MX_temp_exp_TOR_item_COICOP_4dig.dta
  $survey/waste/MX_all_lines_no_COICOP.dta

- MZ_2009_IOF.do
  Take raw household survey data from Mozambique and harmonize module to produce clean final working data file.

  Inputs:
  $survey/raw/Mozambique2009/agregado_v3.dta
  $survey/raw/Mozambique2009/controle_v3.dta
  $survey/raw/Mozambique2009/iof_ponderadores3.dta
  $survey/raw/Mozambique2009/despesas_diarias_v3.dta
  $survey/raw/Mozambique2009/despesas_mensais_v3.dta
  $survey/raw/Mozambique2009/bens_duraveis_v3.dta
  $survey/raw/Mozambique2009/autoconsumo_v3.dta
  $survey/raw/Mozambique2009/receita_v3.dta
  $survey/raw/Mozambique2009/verificacao_de_produtos_v3.dta
  $survey/raw/additional/COICOP_label_2dig.dta

$survey/raw/additional/COICOP_label_3dig.dta
$survey/raw/additional/COICOP_label_4dig.dta

Outputs:

$survey/waste/MZ_household_cov_original.dta
$survey/waste/MZ_all_lines_raw_no_TOR.dta
$survey/waste/MZ_all_lines_raw.dta
$survey/proc/Mozambique2009/MZ_household_cov.dta
$survey/proc/Mozambique2009/Mozambique2009_TOR_stats_for_crosswalk.xls
$survey/proc/Mozambique2009/MZ_exp_full_db.dta
$survey/proc/Mozambique2009/MZ_temp_exp_TOR_item_COICOP_4dig.dta

---

- NE_2007_ENBCM.do

  Take raw household survey data from Niger and harmonize module to produce clean final working data file.

  Inputs:

  $survey/raw/Niger2007/r02.dta
  $survey/raw/Niger2007/menageidentifiant.dta
  $survey/raw/Niger2007/r01.dta
  $survey/raw/Niger2007/r15.dta
  $survey/raw/Niger2007/enbc2007pivot.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/NE_household_cov_original.dta
  $survey/waste/NE_all_lines_raw.dta
  $survey/proc/Niger2007/NE_household_cov.dta
  $survey/proc/Niger2007/Niger2007_TOR_stats_for_crosswalk.xls
  $survey/proc/Niger2007/NE_exp_full_db.dta
  $survey/proc/Niger2007/NE_temp_exp_TOR_item_COICOP_4dig.dta

---

- PE_2017_ENAHO.do

  Take raw household survey data from Peru and harmonize module to produce clean final working data file.

  Inputs:

  $survey/raw/Peru2017/enaho01-2017-200.dta
  $survey/raw/Peru2017/enaho01a-2017-300.dta
  $survey/raw/Peru2017/enaho01-2017-100.dta
  $survey/raw/Peru2017/enaho01-2017-601.dta
  $survey/raw/Peru2017/enaho01-2017-602.dta
  $survey/raw/Peru2017/enaho01-2017-603.dta
  $survey/raw/Peru2017/enaho01-2017-604.dta

$survey/raw/Peru2017/enaho01-2017-605.dta
$survey/raw/Peru2017/enaho01-2017-606.dta
$survey/raw/Peru2017/enaho01-2017-606d.dta
$survey/raw/Peru2017/enaho01-2017-607.dta
$survey/raw/Peru2017/enaho01-2017-610.dta
$survey/raw/Peru2017/enaho01-2017-611.dta
$survey/raw/Peru2017/enaho01-2017-612.dta
$survey/raw/Peru2017/enaho01a-2017-300.dta
$survey/raw/Peru2017/enaho01a-2017-300.dta
$survey/raw/Peru2017/enaho01a-2017-400.dta
$survey/raw/Peru2017/enaho01-2017-100.dta
$survey/raw/Peru2017/enaho01-2017-100.dta
$survey/raw/Peru2017/PE_ENAHO_crosswalk_TOR.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_1.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_3.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_4.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_7.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_8.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_9.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_10.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_11.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_12.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_13.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_16.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_17.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_18.dta
$survey/raw/Peru2017/PE_COICOP_crosswalk_mod_78.dta
$survey/raw/Peru2017/PE_ENAHO_crosswalk_product_code.xls
$survey/raw/additional/COICOP_label_2dig.dta

Outputs:

$survey/waste/PE_household_cov_original.dta
$survey/waste/PE_all_modules_appended.dta
$survey/waste/PE_all_lines_raw.dta
$survey/waste/PE_all_lines_COICOP_crosswalk.dta
$survey/proc/Peru2017/PE_household_cov.dta
$survey/proc/Peru2017/Peru2017_TOR_stats_for_crosswalk.xls
$survey/proc/Peru2017/PE_exp_full_db.dta
$survey/proc/Peru2017/PE_temp_exp_TOR_item_product_code_4dig.dta

---

- PH_2018_FIES.do

  Take raw household survey data from Philippines and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Philippines2018/2018_FIES_vol2.csv

Outputs:

$survey/raw/Philippines2018/2018_FIES_vol2.csv
$survey/waste/PH_household_cov_original.dta
$survey/proc/Philippines2018/PH_temp_exp.dta

- PK_2019_PSLM.do

  Take raw household survey data from Pakistan and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Pakistan2019/plist.dta
  $survey/raw/Pakistan2019/sec_5a.dta
  $survey/raw/Pakistan2019/sec_2ab.dta
  $survey/raw/Pakistan2019/sec_6a.dta
  $survey/raw/Pakistan2019/sec_12ce.dta
  $survey/raw/Pakistan2019/sec_12a.dta
  $survey/raw/Pakistan2019/ppp conversion factor.dta

  Outputs:

  $survey/waste/PK_household_cov_original.dta
  $survey/proc/Pakistan2019/PK_temp_exp.dta

- PY_2011_EIGyCV.do

  Take raw household survey data from Paraguay and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Paraguay2011/Datos de la poblacion reg02t.dta
  $survey/raw/Paraguay2011/Datos de la vivienda y del Hogar reg01.dta
  $survey/raw/Paraguay2011/Agregado a nivel de transaccion agregado.dta

  Outputs:

  $survey/waste/PY_household_cov_original.dta
  $survey/waste/PY_all_lines_raw.dta
  $survey/proc/Paraguay2011/PY_household_cov.dta
  $survey/proc/Paraguay2011/Paraguay2011_TOR_stats_for_crosswalk.xls
  $survey/proc/Paraguay2011/PY_exp_full_db.dta
  $survey/proc/Paraguay2011/PY_temp_exp_TOR_item_product_code_4dig.dta

---

- RW_2014_EICV.do

  Take raw household survey data from Rwanda and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Rwanda2014/cs_s1_s2_s3_s4_s6a_s6e_s6f_person.dta
  $survey/raw/Rwanda2014/cs_s8b_expenditure.dta

$survey/raw/Rwanda2014/cs_s8a3_expenditure.dta
$survey/raw/Rwanda2014/cs_s8a2_expenditure.dta
$survey/raw/Rwanda2014/cs_s8a1_expenditure.dta
$survey/raw/Rwanda2014/cs_s8c_farming.dta
$survey/raw/Rwanda2014/cs_s0_s5_household.dta
$survey/raw/Rwanda2014/cs_s1_s2_s3_s4_s6a_s6e_s6f_person.dta
$survey/raw/Rwanda2014/cs_s6b_employement_6c_salaried_s6d_business.dta
$survey/raw/Rwanda2014/cs_s9b_transfers_in.dta
$survey/raw/Rwanda2014/cs_s10b_goods.dta
$survey/raw/additional/COICOP_label_2dig.dta
$survey/raw/additional/COICOP_label_3dig.dta
$survey/raw/additional/COICOP_label_4dig.dta

Outputs:

$survey/waste/RW_household_cov_original.dta
$survey/waste/RW_all_lines_raw.dta
$survey/proc/Rwanda2014/RW_household_cov.dta
$survey/proc/Rwanda2014/Rwanda2014_TOR_stats_for_crosswalk.xls
$survey/proc/Rwanda2014/RW_exp_full_db.dta
$survey/proc/Rwanda2014/RW_temp_exp_TOR_item_COICOP_4dig.dta

- SN_2008_EDMC.do

  Take raw household survey data from Senegal and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Senegal_Dakar2008/CQ02Aan.dta
  $survey/raw/Senegal_Dakar2008/CQ06Aan.dta
  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/SN_household_cov_original.dta
  $survey/waste/SN_all_lines_raw.dta
  $survey/proc/Senegal_Dakar2008/SN_household_cov.dta
  $survey/proc/Senegal_Dakar2008/Senegal_Dakar2008_TOR_stats_for_crosswalk.xls
  $survey/proc/Senegal_Dakar2008/SN_exp_full_db.dta
  $survey/proc/Senegal_Dakar2008/SN_temp_exp_TOR_item_COICOP_4dig.dta

---

- TN_2010_EBC.do

  Take raw household survey data from Tunisia and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/Tunisia2010/bdg2010T2.dta

$survey/raw/Tunisia2010/produit.dta
$survey/raw/Tunisia2010/menage.dta
$survey/raw/additional/COICOP_label_2dig.dta
$survey/raw/additional/COICOP_label_3dig.dta
$survey/raw/additional/COICOP_label_4dig.dta

Outputs:

$survey/waste/TN_household_cov_original.dta
$survey/waste/TN_all_lines_raw.dta
$survey/raw/Tunisia2010/TN_household_cov.dta
$survey/raw/Tunisia2010/Tunisia2010_TOR_stats_for_crosswalk.xls
$survey/raw/Tunisia2010/TN_exp_full_db.dta
$survey/raw/Tunisia2010/TN_temp_exp_TOR_item_COICOP_4dig.dta

- TN_2010_EBC.do

  Take raw household survey data from Tanzania and harmonize expenditure module to produce clean final working data file.

  Inputs:

  $survey/raw/additional/COICOP_label_2dig.dta
  $survey/raw/additional/COICOP_label_3dig.dta
  $survey/raw/additional/COICOP_label_4dig.dta

  Outputs:

  $survey/waste/TZ_household_cov_original.dta
  $survey/waste/TZ_all_lines_raw.dta
  $survey/raw/Tanzania2012/TZ_household_cov.dta
  $survey/raw/Tanzania2012/Tanzania2012_TOR_stats_for_crosswalk.xls
  $survey/raw/Tanzania2012/TZ_exp_full_db.dta
  $survey/raw/Tanzania2012/TZ_temp_exp_TOR_item_COICOP_4dig.dta
  $survey/waste/SECTA1_to_merge.dta
  $survey/waste/SECTA2_to_merge.dta
  $survey/waste/TZ_all_lines_diary_assign_TOR.dta
  $survey/waste/TZ_all_lines_diary_assign_TOR_autre.dta