**Readme Replication of “Consumption Tax, Informality and Redistribution”**

* The paper is built with STATA and its code consists of dofiles
* This paper contains two key sources of data detailed in part (1) and (2) and some auxiliary data detailed in part (3).

(1) **Micro level household survey data** **of 32 countries listed below.** 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. We list below the country names, survey acronyms and years for anyone interested in replicating our analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **Country name** | **Survey** | **Year** | **Source** |
| Benin | EMICOV | 2015 | Stat. Office |
| Bolivia | ECH | 2004 | Stat. Office |
| Brazil | POF | 2009 | Stat. Office |
| Burkina Faso | EICVM | 2009 | Stat. Office |
| Burundi | ECVM | 2014 | World Bank |
| Cameroon | ECAM | 2014 | World Bank |
| Chad | ECOSIT | 2003 | World Bank |
| Chile | EPF | 2017 | Stat. Office |
| Colombia | ENIG | 2007 | Stat. Office |
| Comoros | EDMC | 2013 | Stat. Office |
| Congo DRC | E123 | 2005 | World Bank |
| Congo Rep | ECOM | 2005 | World Bank |
| Costa Rica | ENIGH | 2014 | Stat. Office |
| Dominican Rep | ENIGH | 2007 | Stat. Office |
| Ecuador | ENIGHUR | 2012 | World Bank |
| Eswatini | HIES | 2010 | World Bank |
| Mexico | ENIGH | 2014 | Stat. Office |
| Montenegro | HBS | 2009 | World Bank |
| Morocco | ENCDM | 2001 | World Bank |
| Mozambique | IOF | 2009 | World Bank |
| Niger | ENCBM | 2007 | World Bank |
| Papua New Guinea | HIES | 2010 | World Bank |
| Peru | ENAHO | 2017 | Stat. Office |
| Rwanda | EICV | 2014 | World Bank |
| SaoTome | IOF | 2010 | World Bank |
| Senegal | EDMC | 2008 | World Bank |
| Serbia | HBS | 2015 | World Bank |
| South Africa | IES | 2011 | U. of Cape Town |
| Tanzania | HBS | 2012 | World Bank |
| Tunisia | ENBCNV | 2010 | Stat. Office |
| Uruguay | ENIGH | 2005 | Stat. Office |

While the microdata cannot be provided, we provide all the meta-data from their analysis, and a full replication of all core figures and tables which are not directly run on the micro-data.

The **master.do** file details all inputs used and outputs produced to allows for full replication.

To harmonize each of these micro data files, we provide the 32 do files, whose name are

*CountryISO\_year\_surveyname* (e.g. for *Mexico MX\_2014\_ENIGH.do*). These then produce harmonize datasets of the survey’s expenditure module including the place of purchase categorized for each purchase, which are used to produce the figures and tables in the paper. We provide the raw data for one survey (Mexico ENIGH 2014) as an example, since it can be downloaded online on open access. The data and corresponding dofiles are contained in the folder “**Surveys**”.

(2) **Mexican micro level price quotes merged with establishment census data**

Accessed from INEGI’s microdata lab**.** The confidential data from the micro data-lab was used to analyze the Mexican VAT reform (Section 6 of the paper).

Detailed at length as part of **Readme\_MX.docx**

INEGI is the National Institute of Statistics and Geography (INEGI) in Mexico. Access to the microdata lab may be obtained upon registering with INEGI. Access to the microdata lab is only granted for public policy and academic research purposes. Researchers interested in access to the data may contact [microdatos@inegi.org.mx](mailto:microdatos@inegi.org.mx) . More information on process to access can also be found here: <https://www.inegi.org.mx/app/microdatos/solicitud/>

The process time that it takes to register with INEGI may vary depending on whether the university that the researcher is affiliated with has a pre-existing agreement with INEGI. If a pre-existing agreement is in place, registration likely takes one to two months. In case of no pre-existing university or institute agreement, the interested user needs to directly sign an agreement with INEGI, and this may take several months to process.

Access to the microdata must be done on-site at INEGI in Mexico City. There is also an option for remote execution of scripts and do files. This service was implemented during the global health pandemic, and it is active at the time of this README (May 2023) though there is no official policy as to whether it will remain active on the future. With the remote service, the researcher sends scripts via email; staff at INEGI execute the scripts. There is no option of remote access to the data. For more information on the dates of the next training please reach out to: [microdatos@inegi.org.mx](mailto:microdatos@inegi.org.mx)

**(3) External Supporting datasets**

**World Development Indicators**

Data on GDP per capita (US$ and LCU), PPP Conversion factor...

The exact dataset we downloaded on 12/17/2019 is available here: https://databank.worldbank.org/embed/GDP/id/7ca22941

see Excel: Country\_information.xlsx

**Findex Data**

Input:Findex17\_SavingsData.dta

Share of people with positive savings at the country\*decile level

Obtained by email on 18th of January 2021 from the Findex team. Aggregated data are publicly avaible here (https://globalfindex.worldbank.org/index.php/#data\_sec\_focus).

We had to email the FINDEX team to obtain this data at the decile level.

see dofile: Savings\_from\_Findex.do

**Enterprise and Informal Surveys**

Firm level data on registration status, total sales and sector. Publicly available on the Enterprise Survey website: https://login.enterprisesurveys.org/content/sites/financeandprivatesector/en/library.html

We also had to email the Enterprise Survey team to access the few Enterprise and Informal surveys that were conducted before 2005 (Bangladesh, Brazil, India, Pakistan, Senegal, Tanzania, Uganda)

see dofiles: WBES\_data\_creation.do, WBES\_analysis.do

**Passport Database on Modern and Traditional Stores**

Sales and number of establishments of traditional and modern stores in each country of our sample

Obtained through the Euromonitor website on 18/11/2020 (https://go.euromonitor.com/passport.html), access requires a subscription

see Excel: Passport\_Stats\_22-12-2020\_2010\_GMT.xls

see dofile: country\_specific\_formal\_assignment.do

**Data on VAT Thresholds**

The value of the VAT exemption threshold in each country of our main sample (lack Paraguay!)

Imported from this Excel file: Retail\_VAT.xls. The sources of this data are detailed in the Excel. For most of the countries the data were downloaded in January 2021 from the International Bureau of Fiscal Documentation country-reports (https://research.ibfd.org/#/).

see dofile: retail\_report.do

**Data on Personal Income Taxes and their thresholds**

Data obtained from Anders Jensen and part of the AER (2022) paper: https://www.aeaweb.org/articles?id=10.1257/aer.20191528

data "$main/data/PIT\_parameters\_AJ.dta"