

Income Taxation in Latin America and the Caribbean: Constructing a New Dataset Based on Household Survey Datasets

RG-K1198

1. Background and Justification

Most studies analyzing the macroeconomic determinants and implications of tax policy for the developing world have typically conducted their analysis at the country level (e.g., Végh and Vuletin, 2015, and Gunter, Riera-Crichton, Végh and Vuletin, 2017). Moreover, given the lack of readily available data regarding relevant tax rates, tax brackets, deductions, and tax coverage, most studies focusing on a large group of countries have typically been unable to engage in a proper measurement of the personal income tax. For example, Végh and Vuletin (2015), who analyze how tax policy is conducted over the business cycle for 62 countries for the period 1960-2013, focus their attention solely on the highest personal income tax rate (as opposed to properly measuring the average marginal income tax rate).

Considering these data-driven limitations, this research project's main objective is to build a novel income tax dataset for the LAC region using proper tax code information and Household Survey (HS) Datasets (e.g., Encuesta Permanente de Hogares in Argentina). Moreover, considering that these Household Survey Datasets are conducted for several urban centers (and sometimes rural areas), typically located in different subnational units (e.g., provinces, departments, and states), this research project aims at building a novel income tax dataset at the national level as well as at the urban centers/rural areas level (e.g., Gran Buenos Aires, La Plata, Mendoza, Cordoba and La Rioja, among others, in Argentina).

The Research Network project "Income Taxation in Latin American and the Caribbean: Constructing a New Dataset Based on Household Survey Datasets" will fund proposals from Latin American and Caribbean research centers that are willing to participate in the construction of the abovementioned dataset. We plan to commission a maximum of 11 national databases in total. The exact number of selected proposals will depend on the quality of the proposals received and the budget available. Submission of proposals should be in tune with the sections described below. Once completed, this dataset will be used for the creation of knowledge products including working papers regarding the macroeconomic determinants and implications of income tax policy.

2. Objective

As described in the previous section, the main objective of this research project is to build a novel income tax dataset for the LAC region at the national level as well as at the urban centers/rural areas level. Since most countries in LAC have been collecting HS Data for several decades now, this research project is expected to have a large coverage of LAC countries and time span (typically annual frequency, and quarterly when available).

Income Taxation in Latin America and the Caribbean: Constructing a New Dataset Based on Household Survey Datasets

3. Scope, Methodology, and Data

This project aims at building a novel dataset that helps to understand, among many other important tax issues, the macroeconomic determinants and implications of personal income tax policy in the LAC region.

Specific variables that each proposal will measure, for this database-driven project, are described below.

3. a. General consideration

- The main objective of this database is to gather a wide number of variables regarding taxation and income distribution in LAC.
- For this purpose, most of those variables must be generated using microdata from HS.
- The database must be built at the regional level (e.g., provinces, departments, subnational states) in order to exploit geographic heterogeneity within each country.
- Time coverage must be as long as possible, given HS availability.
- Data frequency would be annual, semiannual and/or quarterly depending on the periodicity at which HS are published.

3. b. Tax data

Tax data collection consists of gathering key variables regarding personal income tax (PIT) and payroll tax. In order to compile such data, standard tax incidence analysis must be computed for each country.

The unit of analysis must be the individual, and the welfare indicator must be household per capita income.

In order to properly calculate income taxes, it is essential to (i) specify how both taxes are calculated (e.g., tax rates, tax brackets, deductions, and tax base/coverage in PIT) and (ii) specify translation assumptions. Once tax incidence is computed, the following variables must be reported at the subnational level (e.g., provinces, departments, subnational states) as well as at the national level.

Considering typical well-known sub-declaration of incomes in HS, and in order to calculate tax incidences and relevant incomes, the literature has adopted two approaches. One approach is simply not to adjust reported incomes in HS. Another approach is to adjust (somewhat) for under-declaration of incomes. This project calls for taking the first approach; that is to say, we would like to have all subsequent variables be calculated without making any adjustment. Having said that, we would appreciate if the research team has ideas on how implement income adjustments to be discussed at our seminar in October at IDB Headquarters.

Income Taxation in Latin America and the Caribbean: Constructing a New Dataset Based on Household Survey Datasets

i. Personal Income Tax (PIT)

- **Effective PIT rate.** Once those who pay PIT are identified and how much they pay is determined, the effective tax rate is computed considering the total amount of payers in each PIT bracket.
- **PIT burden by percentile**
- **Total PIT burden**
- **PIT payers by percentile**
- **Brackets, rates, and number of total deductions of PIT**

ii. Labor Tax (Payroll)

- **Legal tax rate (for employer and for employee)**
- **Payroll tax burden by percentile**
- **Total Payroll burden**
- **Payroll tax payers by percentile**

3. c. Non-Tax data

Non-Tax data collection consists of relevant variables regarding income, employment status, type of employment, informality, sources of income, and income distribution of each subnational unit (e.g., provinces, departments, subnational states) as well as at the national level.

i. Income variables

- **Aggregate Gross Total Income.** Sum of income received by each individual (e.g., salaries, pensions, rents, dividends, etc.)
- **Net earnings.** Aggregate Gross Income less government transfers and dividends, interest and rents.
- **Aggregate Gross Government-driven Income.** Sum of salaries received by individuals whose main occupation is public employment in the government sector.
- **Aggregate Gross Private-driven Income.** Aggregate Gross Total Income less Aggregate Gross Government-driven Income.
- **Net Government-driven earnings.** Sum of salaries received by individuals whose main occupation is public employment in the government sector less government transfers and dividends, interest and rents.
- **Net Private-driven earnings.** Net earnings less Net Government-driven earnings

ii. Employment variables

- **Total population**

Income Taxation in Latin America and the Caribbean: Constructing a New Dataset Based on Household Survey Datasets

- **Total employment**
- **Total unemployment**
- **Labor force**
- **Labor force participation**
- **Population by age (less than 15 years, between 16-65, more than 65)**
- **Total private employment**
- **Total public employment**
- **Total wages:** sum of wages.
- **Total private wages:** sum of wages for those individuals declaring main occupation as private.
- **Total public wages:** sum of wages for those individuals declaring main occupation as government public.
- **Labor informality**
- **Number of people employed**
- **Average hours worked (all workers)**
- **Average hours worked (private workers)**
- **Average hours worked (public workers)**
- **Average years of education for private workers**
- **Average years of education for public workers**
- **Average years of education for informal workers**

iii. Income Distribution and Inequality variables

- **Pre (PIT and payroll) tax income by percentile**
- **Post (PIT and payroll) tax income by percentile**
- **Pre (PIT and payroll) tax Gini Coefficient**
- **Post (income and payroll) tax Gini Coefficient**
- **Kakwani index for PIT**
- **Kakwani index for payroll**
- **Poverty rate (using US\$ 1.90 and US\$ 4)**
- **Does the HS include urban centers'/rural areas' own Consumer Price Indices?**

3. d. Complementary material

The database, containing the above variables, should be accompanied by the following three additional documents in order to support not only the calculation of the variables but also the understanding of how taxes work in each country and how they have been modified over time.

- **Wide description of both taxes (PIT and payroll)**
- **Wide description of changes in both taxes during the period covered by the data.** Not only tax rate changes, but also changes in brackets, deductions, tax base, etc. Raw data on these narratives (e.g., laws, newspapers articles, etc.) would be greatly appreciated.

Income Taxation in Latin America and the Caribbean: Constructing a New Dataset Based on Household Survey Datasets

- **Methodological Report.** A document that details how the incidence of each tax was computed, with all assumptions used. It should contain an analysis of the incidence results and how they have been altered before each tax change.

3. e. Additional data

Household Surveys data and Stata .do files must be delivered in order to replicate the results.

4. Selection Criteria

Research institutions only (including think-tanks) may present proposals. The IDB seeks to produce up to eleven (11) national databases. The final number of proposals accepted will depend on the quality of the proposals received and the available budget. Proposed budgets will be evaluated in accordance with the scope of the work proposed.

The research proposals will be selected based on the following criteria:

- a) 40 percent based on clarity and quality in the description on the data availability and analysis to be conducted;
- b) 30 percent based on greater temporal coverage of data;
- c) 30 percent based on experience of the research team.

To be considered for the project, the research proposal should include the following:

- a) A brief survey of previous studies on this topic conducted on the country or countries;
- b) Description of data availability (both in terms of urban centers/rural areas as well as time coverage for alternative time frequencies);
- c) A detailed description of the strategy proposed for collecting and constructing each relevant variable (see Section 3 for details);
- d) Indicate main country of interest to build the previously described dataset. Also indicate if the research team would be interested (given the expertise in income tax and household data knowledge) in potentially working with another country. This element may be important in case of receiving highly competitive proposals for the same country.

In all cases, the final dataset employed in the analysis will be provided to the IDB along with Excel and .dta files and a .do file, together with all necessary "raw" Household Survey Dataset (in .dta) files to replicate final data, as well as a short companion .doc document describing the data and methodology. An exception may be made in the case of proprietary data, in which case the team must provide the appropriate documentation.

Income Taxation in Latin America and the Caribbean: Constructing a New Dataset Based on Household Survey Datasets

5. Proposal submission

Research institutions interested should submit a proposal no later than **July 28, 2017** by e-mail to Elton Mancilla at red@iadb.org. The following information is required for submitting the proposal:

- The name of the research leader and a list of other researchers involved. The center should demonstrate the ability of their research team to meet the objectives of the project, including relevance of prior experience. Curricula vitae of all researchers involved in the whole project may appear in a separate annex. Subsequent substitutions for researchers originally specified in the proposal may be made with prior approval from project coordinators, but the research leader must lead the entire project until its full completion.
- A budget (in a separate annex) indicating the time and resources that will be used within the context of the research work plan. The budget proposed by the research center should disaggregate items financed by the IDB contribution and those financed by the research center. The budget should distinguish among amounts assigned to professional honoraria, “overhead” and other major categories of research expenditures. **The proposal and corresponding budget *must* be sent in separate files.**
- Institutions must provide the name and contact information of their legal representative, with authority to sign contracts with the IDB, if selected to conduct the study.

Proposing **research institutions** should be registered as Research Network members (contact Elton Mancilla at red@iadb.org for this purpose) and should be based in the Latin American and Caribbean region. U.S. and European institutions do not qualify as members of the Research Network. However, researchers from the United States and Europe can participate with research teams from Latin American and Caribbean proposing institutions.

6. Coordination

The project will be coordinated by Leonardo Gasparini (CEDLAS), Jorge Puig (CEDLAS), and Guillermo Vuletin (IDB/RES).

The tentative schedule of activities is as follows:

- **July 28, 2017:** Due date for **receiving proposals**. Institutions should ensure that the complete documentation is submitted no later than **July 28, 2017** by e-mail to Elton Mancilla at red@iadb.org. Complete documentation includes the following components: registration form with all information requested, the research proposal, budget, and curriculum vitae (CVs up to three pages long).

Income Taxation in Latin America and the Caribbean: Constructing a New Dataset Based on Household Survey Datasets

- **August 2, 2017:** Announcement of **selected research proposals**. After this announcement, research teams will start working on the creation of the dataset interacting with the Research Network's coordinators, and in preparation for a seminar to be held in October at IDB Headquarters of IDB in Washington DC.
- **September 20, 2017:** Due date for receiving a **first draft** of dataset and companion document describing the data and methodology.
- **October 2, 2017: Discussion seminar** in Washington, D.C., with the Research Leaders of each team for the purposes of presenting and discussing dataset created by research teams as well as technical aspects. At this point, research teams will present a very polished version of the dataset created. The main purpose of this seminar is to discuss, brainstorm, and finish making decisions in order to have a fairly homogeneous dataset across different countries.
- **December 15, 2017:** Deadline for a **second version** of the dataset and companion document describing the data and methodology.
- **February 15, 2017:** Deadline for a **final version** of the dataset and companion document describing the data and methodology. Data must be submitted by this date (no exceptions).

7. Financial Contribution and Payment Schedule

The IDB will contribute up to US\$17,000 (or domestic currency equivalent) for each study depending on the scope of work proposed. The payment schedule is as follows:

- **20 percent** within 30 days of signing the formal agreement between the IDB and the respective research center.
- **30 percent** within 30 days of receiving and approving the first draft of the dataset and short companion document describing the data and methodology.
- **30 percent** within 30 days of receiving and approving the second draft of the dataset and short companion document describing the data and methodology.
- **20 percent** within 30 days of receiving and approving the final version of dataset and short companion document describing the data and methodology.