

Reproducibility Package README

The Worldwide Bureaucracy Indicators (WWBI) – Version 3.0 – Subsample for Brazil - 2015 and Wage Bill Indicators

Overview

The WWBI are a unique cross-national dataset on public sector employment and wages that aims to help researchers, development practitioners, and policymakers gain a better understanding of the personnel dimensions of state capability, the footprint of the public sector within the overall labor market, and the fiscal implications of the public sector wage bill. The dataset is derived from nationally representative household surveys augmented with administrative data, thereby complementing existing, expert perception-based approaches. It covers three primary categories of variables: public-sector employment characteristics, corresponding wages and compensations, and wage bill. Together, these indicators provide uniform and reliable estimates across different periods and geographical areas, ensuring comparability and consistency.

The dataset initially debuted in 2019 and has undergone annual updates since its inception. In 2022, version 3.0 was introduced, marking a significant milestone. This update notably extended the dataset's coverage to encompass a broader array of countries, offering a more comprehensive representation. Additionally, it substantially augmented the number of indicators integrated within the dataset, enhancing its depth and scope.

Due to the extensive nature of the input data derived from official household surveys and the expansive scale of the current WWBI dataset, encompassing a substantial total of 1,067 surveys, it is regrettable that sharing the entirety of this data and performing computations on a standard computer is impractical. For the sake of reproducibility, this version of the program operates on a subset of data specific to Brazil and for the year 2015, and for the wage bill indicators for all countries in years 2000-2020.

This comprehensive reproducibility package is designed as an essential resource for policymakers, quantitative and comparative researchers intrigued by cross-national and temporal variations in public sector size and structure and the broader community interested in scrutinizing, verifying, and expanding upon the discoveries of the dataset. Our intention is to provide an intricate overview of the data, code, and procedures utilized in the construction of the dataset. Through this transparency, we strive to encourage deeper comprehension and active involvement in the crucial discourse concerning public sector employment and compensation practices.

Data

The most important data source for the WWBI are the World Bank's housing nationally representative surveys encompassing household welfare and labor force surveys across the globe. These are stored under different harmonization projects, such as:

- World Bank International Income Distribution Database (I2D2)
- World Bank Global Monitoring Database (GMD)
- World Bank Global Labor Database (GLD)

- World Bank Latin America and the Caribbean Equity Lab (LABLAC)

These surveys employ a standardized taxonomy, ensuring uniform data categorization and harmonization across countries. WWBI version 3.0 aggregates data from all these sources, comprising 690 surveys sourced from the I2D2 dataset, spanning 118 countries. Additionally, it integrates 15 surveys from 11 countries sourced from the GMD database and 3 surveys from a single country sourced from the GLD database and 12 surveys from the LABLAC initiative.

The dataset also incorporates data from other external sources such as:

- Eurostat European Union Statistics on Income and Living Conditions (EU-SILC)
- LIS Cross-National Data Center, Luxembourg Income Study Database (LIS)
- International Monetary Fund (IMF) Government Compensation and Employment Dataset, 2020
- International Comparison Program (ICP) 2017 Cycle Data for Researchers database

WWBI version 3.0 includes 343 surveys from the EU-SILC project, representing 29 countries and 4 surveys from the LIS dataset. Moreover, the WWBI encompasses the IMF Government Compensation and Employment Dataset, incorporating information on the wage bill as a percentage of total expenditure and as a percentage of GDP. In its entirety, the WWBI comprises a comprehensive collection of 1,067 surveys corresponding to 144 countries and territories worldwide.

Data for the Reproducibility package

This version of the program operates on a subset of data specific to Brazil for the year 2015, and for the wage bill indicators (Wage bill as a percentage of GDP and Wage bill as a percentage of Public Expenditure) for the years 2000-2020. This code snippet is designed to execute the provided do files using input data for reproducing only these results.

The data for Brazil is an amalgamation of harmonized surveys spanning the years 2001-2015, sourced from the World Bank's I2D2 harmonization collection. This dataset is specifically formatted as an input dataset for the country of Brazil. The survey files included in this dataset are the following:

Country	Year	Type	Survey name	Harmonization initiative
Brazil	2001	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2002	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2003	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2004	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2005	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2006	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2007	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2

Brazil	2008	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2009	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2011	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2012	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2014	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2
Brazil	2015	Publicly available	Pesquisa Nacional por Amostra de Domicilios	I2D2

As mentioned earlier, the WWBI primarily depends on official household and labor force survey data. To ensure consistency across comparisons, the program utilizes a harmonized version of this data curated by the I2D2 or other harmonization teams within the World Bank as its input. The harmonized surveys used in the WWBI are exclusive to the World Bank and not accessible outside of it. Consequently, while the official survey data for Brazil is publicly available, the specific version of the data utilized for the WWBI is not externally accessible. However, for reproducibility purposes, we are sharing this subset of data, enabling the creation of WWBI indicators for a specific country.

Additionally, the WWBI dataset features the data for wage compression ratios in public sector, coming from the International Comparison Program (ICP). This data is part of the worldwide statistical initiative to collect and compile comparable price and national accounts expenditure data and estimate purchasing power parities (PPPs) for the world's economies. managed by the ICP Global Office at the World Bank. Indicators from this dataset measure wage compression ratios both within and between public sectors of respective countries, which represent a natural complement to the pay compression ratios between the 90th and 10th percentile of wage earners in the public and private sectors within each country.

Finally, the IMF Government Compensation and Employment Dataset provides the information on the size of the public sector wage bill corresponding to the year 2020 for most countries of the world, except for: Anguilla, Bermuda, Curaçao, Cayman Islands, Montserrat, West Bank and Gaza, Romania, Sint Maarten (Dutch part), Turks and Caicos Islands and Kosovo. The two indicators that come from this dataset are also added to the WWBI dataset. They are "Wage bill as a percentage of GDP" and "Wage bill as a percentage of Public Expenditure".

Getting Started

The first step in this reproducibility package is to open the master do file (0_master.do) and run the code it contains. The file sets up the necessary paths and runs the required Stata code to generate the indicators and summary statistics to create the dataset. The master do file is structured as followed:

Part 0: User inputs and folder paths

This section sets the global main folder path for the reproducibility package and defines various folder paths related to data inputs and globals for variable names and for unit conversions.

Part 1: Sample selection

This section defines temporal coverage of harmonized datasets, runs quality checks of the input data, and defines filters to identify, tag and remove surveys with incomplete data. It also constructs and harmonizes public employment variables and constructs industry and occupational classifications for further analysis.

Part 2: Summary statistics

This section generates all the indicators included in the dataset, including a wide range of metrics such as public sector employment, wage premiums, gender-based wage differentials, indicators evaluating access to benefits and social protection, demographic distribution of employees, gender distribution across various occupations and wage quantiles, as well as educational characteristics of workers.

Part 3: Merging, outlier analysis and exporting

This section transforms the featured indicators into output data, identifies and removes outliers and inconsistent observations and finally, prepares the final dataset to export to excel.

Installation

To reproduce the results in this replicability package, you will need to use the software Stata.

Usage

To successfully run this package, please follow these steps: First, set the user root-path in line 19 of the 0_master.do. Make sure the corresponding data and input files are well saved in the corresponding folder to match the folder roots. Also, make sure you specify the output folder to save the results. Next, run this do file up until the end of it.

Results

The reproducibility package provides a subset of the foundational data and necessary code to reproduce the indicators featured in the Worldwide Bureaucracy Indicators Dataset. To enhance accessibility, this program version concentrates on generating a subset of indicators specifically within the WWBI: Brazil for the year 2015 and the Wage Bill Indicators for the years 2000-2020.

For a comprehensive array of results and indicators encompassing the remaining 144 countries included in the dataset, please refer to:

- For accessing the latest data, Explanatory Note and Codebook → <https://datacatalog.worldbank.org/int/search/dataset/0038132/worldwide-bureaucracy-indicators>
- For accessing the dashboard featuring visualizations based on the latest version of the dataset → <https://www.worldbank.org/en/data/interactive/2019/05/21/worldwide-bureaucracy-indicators-dashboard>
- For accessing reports with key findings of the WWBI dataset → <https://onlinelibrary.wiley.com/doi/pdf/10.1111/puar.13355>

<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099062623201128026/p16870312f991025147b814345180681d74af9845bed>

Contributing

We enthusiastically invite and encourage contributions to the WWBI and to this reproducibility project. Whether you aim to refine or enrich this reproducibility package, your involvement can take various forms—identifying and resolving bugs, proposing innovative features, or suggesting enhancements.

Your contributions are pivotal in elevating the quality and usefulness of this dataset. We deeply value your support in advancing this project to greater heights.

Acknowledgments

The WWBI is a collaboration between the Governance Global Practice and the Development Impact Evaluation unit of the Development Economics Vice Presidency. The original and updated version of the dataset and explanatory note were produced by a core team consisting of Muhammad Faisal Ali Baig, Zahid Hasnain, Daniel Rogger and David Newhouse, with research support from Camilo Bohorquez-Penuela, Vanessa Cheng, Tim Gindling, Aditi Mishra, Turkan Mukhtarova, Flavia Sacco Capurro, Rong Shi, Pablo Suarez-Becerra and Danny Walker. The work was conducted under the overall guidance of James Brumby, Asli Demirguc-Kunt, Shantayanan Devarajan, Asmeen Khan, Tracey Lane, Arianna Legovini, Edward Olowo-Okere and Deborah Wetzel.

Contact

For any inquiries or assistance concerning this reproducibility package or to address any questions you might have regarding the contents and usage of this package, please don't hesitate to contact Flavia Sacco Capurro at fsaccocapurro@worldbank.org. Your feedback and inquiries are highly appreciated.