Worldwide Bureaucracy Indicators (WWBI)

Codebook and Explanatory Note

Bureaucracy Lab

WORLD BANK GROUP

WORLDWIDE BUREAUCRACY INDICATORS

Version 3.1

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, Junying Tong, 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.

The official full name of this dataset is 'Worldwide Bureaucracy Indicators dataset'.

The official abbreviation of this dataset is WWBI.

The current version of the dataset is 3.1 (August 2024)

Always include the version number in analyses using the dataset.

When using this data, please always cite:

World Bank. 2024. Worldwide Bureaucracy Indicators version 3.1

When appropriate, also cite this codebook:

World Bank, 2021. "Worldwide Bureaucracy Indicators version 3.1: Codebook and Explanatory Note", Washington, DC: World Bank.

Enquiries on the WWBI or derivation should be addressed to Zahid Hasnain (zhasnain@worldbank.org) or Faisal Baig (mbaig5@worldbank.org).

WORLD BANK GROUP
Governance

i2i DIME
TRANSFORM DEVELOPMENT

Table of Contents

1. Introduction

The Worldwide Bureaucracy Indicators (WWBI) are a project of the Bureaucracy Lab at the World Bank. The Lab seeks to promote the use of evidence and technology in the study of the civil service through the creation of new datasets, diagnostic instruments, and knowledge products to improve World Bank policy advice and operations. The WWBI are a unique cross-national dataset on public sector employment and wages that aims to fill an information gap, thereby helping 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.

The WWBI encompass three categories of variables: the characteristics of public-sector employment, their wages and compensations, and the overall wage bill. Together, these provide an important, albeit narrow, picture of the skills and incentives of bureaucrats. Indicators on public employment track key demographic characteristics including the size of the public sector workforce (in absolute and relative numbers), their age, and distributions across genders, academic qualifications, and their respective industries and occupations of employment. Variables on compensation capture both the competitiveness of public sector wages (compared to the private sector) as well as wage differentials across sex, education, occupations and industries of employment, and income quantiles between the public and private sectors. The indicators on the size of the wage bill offer a glimpse into the structure and affordability of the public sector within the larger economy. Therefore, the WWBI are primarily directed towards, and are most useful for, quantitative and comparative researchers interested in both cross-national and temporal differences in the size and organization of the public sector. The dataset is constructed with a deliberate effort to harmonize multiple data streams to offer comparable and consistent estimates across time and space.

Version 3.1 of the WWBI enhances both temporal and geographical microdata coverage compared to version 3.0. This update includes 129 new surveys and adds two additional countries to the dataset. Additionally, it updates the data on general government compensation of employees ("wage bill") to the year 2022 for 195 countries in the database (previously updated to 2020 in the 3.0 version). In total, WWBI version 3.1 encompasses 1,072 surveys from all regions of the world.

The WWBI version 3.0 -released in May 2022- made four major improvements on version 2.0. First, it expanded microdata coverage adding 158 new surveys and 9 new countries, setting to 144 the number of countries covered by the database. Second, this version included 110 new indicators. Most of these new indicators correspond to a new addition of the occupational classification of workers, including teachers, medical workers and their corresponding characterization and their relative sizes in the labor market. Also, it adds several indicators on core categories of employment (managers, professionals, technicians, clerks and elementary workers) along with their characterization considering gender and age (detailed in section 7). In addition to the 3 industries identified in version 2.0, this version added a new industrial classification for core public administration, public safety and social security. Third, version 3.0 also included public sector wage premiums along with the compensation patters by sex, and public/private sectors observed in each occupation and among the key industries considered (more information on section 7). Finally, this version updated the data on general government

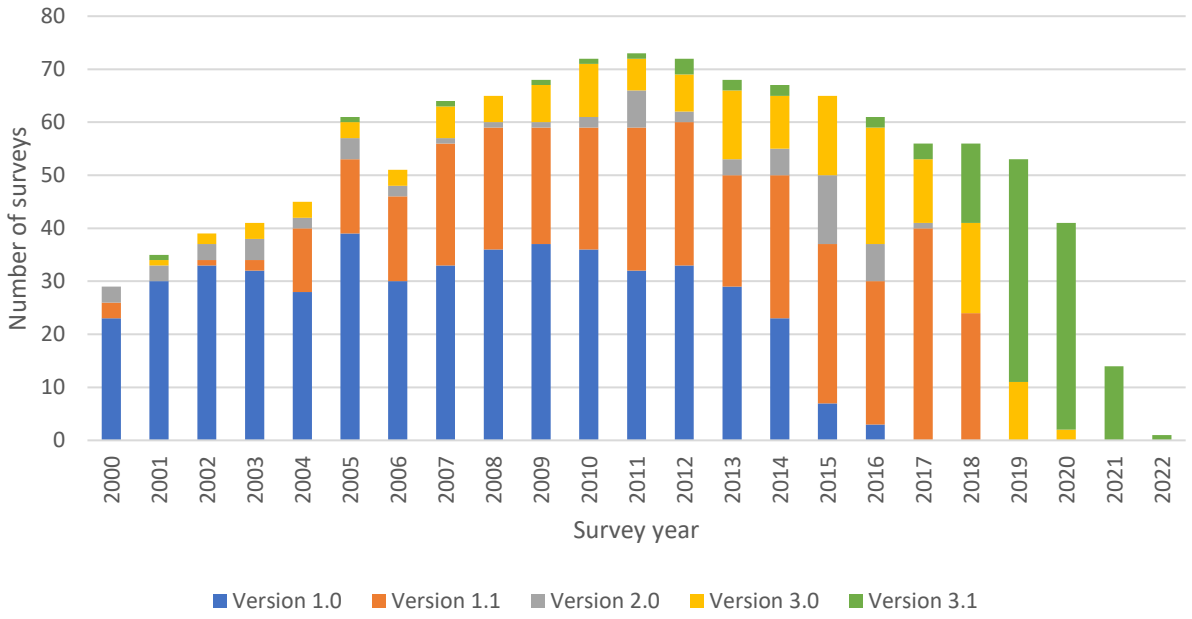
compensation of employees ("wage bill") to year 2020 (2018 in previous version) for 195 of the countries in the database. The previous version (2.0), , released in May 2021 made three improvements on version 1.1. First, it expanded microdata coverage to 909 surveys from 135 countries and territories (up from 846 surveys from 132 countries and territories in version 1.1). Second, it augmented the database coverage to 202 countries (and territories) by including administrative data on compression ratios within the public sector for 167 economics and the relative size of the wage bill data for 177 economies (detailed in section 4). Third, it included 73 new indicators on the decomposition of public sector employment and compensation patterns along three key industries of employment in the public sector; education, healthcare, and public administration (elaborated in detail in section 7.5). Overall, version 2.0 added 100 new indicators to the WWBI dataset and makes total number of indicators 192.

Version 1.1, released in June 2020, extended WWBI's coverage to previously underrepresented regions using new primary data sources, updated temporal coverage to 2018 and made methodological improvements by transforming premiums resulting from wage regressions to their more precise specification (elaborated again in Appendix A3); added additional indicators reflecting the statistical significance (p-values) for these premium regressions; and amended the quality thresholds set out in version 1.0 for the selection of surveys for inclusion (described in section 4 and elaborated in detail in Appendix A1).

To the greatest extent possible, the update is compatible with the earlier version. This note provides a brief guide to the data, their sources and the definitions and transformations used to derive variables included therein. Section 6 and 7 provide detailed descriptions of all variables.

Given the fact that the database is based on harmonized nationally representative household surveys—both welfare and labor force surveys—from around the world, it is possible that for a particular country data may come from different surveys over different years. Every attempt is made to ensure that (if available and meeting the WWBI's data standards) the same surveys are used over years. For countries with multiple surveys available for a given year, the survey from the same source as other years' survey is kept. However, if a unique dataset has a larger number of observations, it is kept in service of greater precision. Although, the indicators generated undergo additional scrutiny to identify outliers and structural breaks, as catalogues in Appendix A2, there still may be inconsistencies in the indicators over time due to these differences in the primary data sources that users will need to consider. A detailed list of countries with their original data source is presented at A4.

Figure 1: WWBI Coverage over years by version release



Source: World Bank. 2024. Worldwide Bureaucracy Indicators version 3.1

2. Definitions

2.1. Public sector

The ability of the WWBI to directly compare public and private sector employment and compensation is only possible with a globally harmonized definition of the public sector. This is hindered by issues of comparability emerging from the heterogenous definition of public employees across countries. In order to circumvent this, the WWBI, as a guiding principle, utilizes the classification of the "public sector" as opposed to the more narrowly defined "general government" as defined by the IMF's Manual on Government Finance Statistics (GFS) and described below and illustrated in Figure 2.

- The *public sector* consists of all institutional units controlled directly, or indirectly, by the central and subnational governments as well as public corporations that are engaged in a market-based activity. The public sector is the general government and public, or state-owned, enterprises. (IMF 2014, 1).

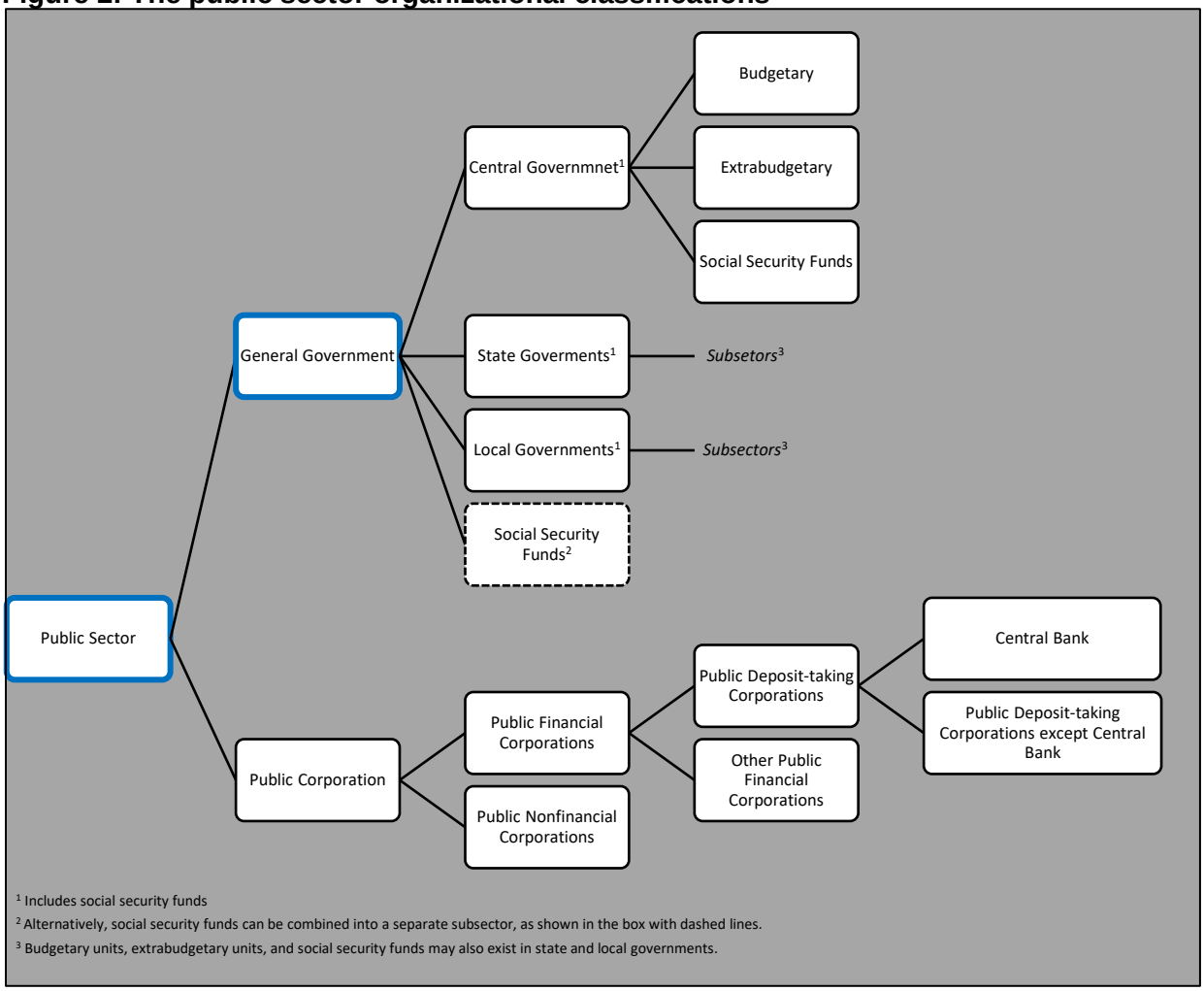
- The *general government* consists of all institutional units in the country that fulfill the functions of government as their primary activity, which includes central and sub-national budget funded and nonmarket, non-profit institutions. (IMF 2016, 1).

Moreover, this definition of public sector employment also corresponds to the one laid out by the International Labor Organization (ILO) within the ILOSTAT database:

> The *total public sector employment* covers all employment of general government sector as defined in System of National Accounts 1993 plus employment of publicly owned enterprises and companies, resident and operating at central, state (or regional) and local

levels of government. It covers all persons employed directly by those institutions, without regard for the particular type of employment contract" (ILOSTAT 2020).

Figure 2: The public sector organizational classifications



Source: IMF (2014)

Utilizing the more broadly defined public sector allows for a more comprehensive comparative analysis between the public and private sectors. Additionally, given the self-reported nature of the household welfare surveys used in the indicators and in order to ensure a globally consistent definition, the public sector as defined in the report includes all individuals employed within core public administration, the security sector, the public sector education and healthcare workforces, individuals employment in public institutions including Central Banks and state owned Enterprises (SOE). Significant effort is made to align all surveys to this broader definition of the public sector.

The only exception to this is for data on countries within the European Economic Area (EEA) between 2004 and 2018. For these countries and years, the analysis is based on the smaller general government aggregate. This was required due to the unique nature of the data and in order to maintain the dataset's adherence to the UN System of National Accounts (SNA), which refers to "public offices at all levels of government, [including] non-market publicly owned hospitals, schools, and social security organizations", but excludes "public or quasi-public

corporations, even when all the equity of such corporations is owned by government units" (European Commission, 2014, 12). Specifically, the classification of public sector employees in the EEA uses economic activity rather than by sector (described in detail in section 6.2).

Box 1: Public Sector in the European Economic Area
The identification of public vs. private sector employees is based on a specific question on the sector of employment for most countries. This is the case for the microdata sourced from the I2D2, LABLAC, and LIS data repositories. The only exception to this is the data sourced from Eurostat's European Union Statistics on Income and Living Conditions (EU-SILC) for countries within the European Economic Area (EEA). For these countries, an approximation of the public sector is made by combining NACE Rev. 2 industry classifications "O" which covers public administration, defense and compulsory social security, P which accounts for human health services and "Q" covering education services to represent the public sector. Therefore, unlike the definition of public sector employees used for surveys sourced from the I2D2, this definition does not include public sector workers employed in public and quasi-public corporations. Therefore, constructing the public sector identifies using industry classification more closely aligns with the definition of the general government as opposed to public sector as defined above in Section 1.2. There are drawbacks associated with this approach as not all individuals employed in education and health services operate in the public sector which may overestimate the size of the public sector. Still, the aggregation of these three provides a fair approximation to the general government, especially for countries in the European Union given the large public sector healthcare and education sectors as is standard practice within the literature (Christofides and Michael 2013; de Castro, Salto, and Steiner 2013; European Commission 2014; Giordano et al. 2015).

2.2. Employment

The classification of employed individual, paid employee, and public paid employee is based on labor and employment status and the type of sector as described below:

- Definitions for total and formal employment are based on the ILO's International Classification of Status in Employment (ICSE), making the WWBI and ILOSTAT databases cross-compatible (fundamental differences in survey coverage, representation, sample size, and timing notwithstanding). According to the ICSE, total employment is defined as:

Persons in employment are defined as all those of working age who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit. They comprise employed persons "at work", i.e., who worked in a job for at least one hour; and

employed persons "not at work" due to temporary absence from a job, or to working-time arrangements (such as shift work, flextime, and compensatory leave for overtime (ILO 2021,5).[1]

2.3. Wages

Wage data in the WWBI denotes the income associated with the occupation of employment used in the analysis (which the individual dedicated most of their time in the week preceding the survey) and excludes both bonuses, allowances, and other in-cash/-kind payments from the same job as well as all additional sources of income (from other jobs) or investments and transfers. Due to the almost complete lack of information on taxes, the wage from primary job is not net of taxes. For all those with self-employment or owners of own businesses, this corresponds to net revenues (net of all costs excluding taxes) or the amount of salary withdrawn from the business.

Certain surveys do include information on work benefits, such as health insurance and social security, but these are not monetized and cannot be added to wages to provide an estimate of total compensation.

Wage information in the surveys is reported in each country's local currency units (LCU), with a diverse array of periodicity. Great care is taken to identify the exact frequency of income for each individual within the surveys and convert all wages to weekly wage after accounting for varying levels of hours worked to ensure credible comparisons across individuals and groups. Additionally, to control for the effect of possibly spurious outliers, the wage variables are winsorized by limiting extreme values in the survey data at the top 0.01 percent level.[2]

3. Data sources

The WWBI data are mainly drawn from primary sources:

- 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)
- Eurostat European Union Statistics on Income and Living Conditions (EU-SILC)
- LIS Cross-National Data Center, Luxembourg Income Study Database (LIS)

[1] The 19th International Conference of Labour Statisticians (ICLS) in 2013 adopted revised standards concerning statistics of work, employment, and labor underutilization. Included within these was a narrowing of the definition of employment to work performed for pay or profit, which would exclude, for example, activities where the self-declared main intended use of the output is for own/family consumption. Hence, these no longer count within employment or towards labor force participation rates.

[2] Winsorization was used in the analysis of wage data in the WWBI to eliminate the impact of extreme values within the dataset on the coefficients to avoid possibly spurious outliers. This approach also ensures that no observations are removed but that outliers above 99.9 percent of the distribution are set to a specified percentile of the data; for example, a 99.9th percentile in the data.

- International Monetary Fund (IMF) Government Compensation and Employment Dataset, 2022 *[Version 3.1 expands coverage from the dataset]*

- International Comparison Program (ICP) 2017 Cycle Data for Researchers database

The main source of data for the WWBI is the World Bank's I2D2 database which stores nationally representative surveys—both household welfare and labor force surveys—globally, harmonizing data using a common taxonomy applied to all countries and surveys. The database contains household and individual data on demographics, employment, and wages. For version 1.0, 470 surveys from 101 countries enumerated between 2000 and 2016 (out of a total sample of 1,394) from I2D2's Revision 6 dataset were used to construct the country-level indicators. To this, 14 additional surveys from OECD nations were added from the LIS Cross-National Data Center in Luxembourg, for a total coverage of 484 surveys and 115 countries (Table 1).

For Version 1.1, two new catalogs were utilized in the development of the indicators; the Eurostat's EUSILC and World Bank LABLAC data catalogs; 343 surveys from 29 countries and 12 surveys from 12 countries, respectively. Since the EU-SILC offers consistent coverage for European countries, 23 surveys used in the version 1.0 were removed in lieu of more consistent measures based on EU-SILC database (10 surveys from the LIS database and 13 from the I2D2 database). Additionally, as briefly discussed in section 1 above and elaborated below in Section 4 and Appendix A1 and A2, 29 additional surveys from the 12D2 were included Worldwide Bureaucracy Indicators (WWBI) in Version 1.1 (compared to version 1.0). Five of these were added since they passed the adjusted quality threshold of 40 percent (compared to the 30 percent employed in Version 1.0) and 24 surveys were added due to the changes in the identification and treatment of outliers. These additions and modifications increased WWBI's coverage to 846 surveys and 132 countries (Table 1). Finally, information on the size of the public sector wage bill were derived from the IMF Government Compensation and Employment Dataset, 2016 and added to the database for years where surveys existed; a total of 1,588 observations.

WWBI version 2.0, used the updated revision 7 of the I2D2 dataset from which 550 surveys are included (75 additional surveys compared to version 1.1) covering 109 countries between 2000 and 2018. To this, 343 additional surveys from 29 EEA member countries were added from the EU-SILC database. Finally, as in version 1.1, 12 surveys from the World Bank's LABLAC and 4 surveys from LIS also make it into the WWBI Version 2.0. This increased the total coverage of WWBI version 2.0 to 909 micro-data surveys from 135 countries .

Version 2.0 also included wage compression ratios in public sector by the International Comparison Program (ICP), available for 167 countries (and territories) and information on the size of the overall wage bill derived from the IMF Government Compensation and Employment Dataset, available for 177 countries (and territories). Adding those increases the geographical coverage to 202 countries (and territories)

Version 3.0 featured 690 surveys from the extended revision 7 of the I2D2 dataset (139 more surveys than version 2.0), covering 118 countries up to 2020. This version also added 15 surveys from 11 countries coming from the GMD database and 3 additional surveys for one country from the GLD database. Additionally, the WWBI 3.0 included the IMF Government Compensation and Employment Dataset with the data updated to 2020 for 195 countries and territories (18 more than in version 2.0).

In version 3.1, 129 new surveys were incorporated from I2D2, GMD, GLD, and the EU-SILC databases, bringing the total to 1,072 surveys. This version also updates the IMF data up to 2022 on Wage Bill for 195 countries. This expansion now covers 202 countries and territories in the dataset.

Table 1: WWBI coverage by data source

WWBI data sources	Version 1.0		Version 1.1		Version 2.0		Version 3.0		Version 3.1	
	surveys	countries	surveys	countries	surveys	countries	surveys	countries	surveys	countries
I2D2	**470**	**104**	**487**	**105**	**550**	**109**	**690**	**118**	**641**	**119**
GMD							**15**	**11**	**16**	**11**
GLD							**3**	**1**	**4**	**1**
LIS	**14**	**14**	**4**	**4**	**4**	**4**	**4**	**4**		
EU-SILC	**-**	**-**	**343**	**29**	**343**	**29**	**343**	**29**	**399**	**30**
LABLAC	**-**	**-**	**12**	**12**	**12**	**12**	**12**	**12**	**12**	**12**
IMF Wage Bill	**358** *(obs)*	**84**	**1,588** *(obs)*	**121**	**6,280** *(obs)*	**177**	**7,331 (*obs*)**	**195**	**7,331 (*obs*)**	**195**
ICP	**-**	**-**	**-**	**-**	**3,497** *(obs)*	**167**	**3,497 (*obs*)**	**167**	**3,497 (*obs*)**	**167**
Total	**484**	**115**	**846**	**132**	**909**	**202**	**1,067**	**202**	**1,072**	**202**

Source: World Bank. 2024. Worldwide Bureaucracy Indicators versions 1.0, 1.1, 2.0, 3.0 and 3.1.

All the indicators included in the WWBI are derived from survey data with the exception of the indicators sourced from the IMP and the ICP which are based on public sector administrative data. While public sector administrative data is potentially a more accurate and detailed measure of employment and wages in the public sector, it does not allow for comparisons with the private sector. Many countries lack administrative and information technology systems to be able to regularly and effectively produce accurate data on public sector employment and compensation. Further, heterogenous adherence to standardized GFS definitions of the public sector or general government, creates challenges for cross-national comparisons.

Household surveys have certain advantages – and some shortcomings - over administrative data as a source of information on public sector and general government employment and wages. One of the main advantages of household surveys is that they provide a rich, consistent, and regularly updated set of variables for a variety of worker characteristics in the public and private sectors that enable robust, controlled comparisons between the two groups. The surveys on which the WWBI relies are some of the most professionally conducted survey undertakings in the world, frequently supported or managed by World Bank or multilateral organization data teams with substantial experience in undertaking such exercises. The harmonization process that brought the surveys used in the WWBI together was managed by experts in surveying from across the World Bank. At the same time, heterogeneity within surveys due to differences in questionnaire design between countries and over the years limits the ability to apply a uniform coding schema to a large set of indicators. Therefore, the WWBI team relies on a core set of variables that are common to most if not all surveys for the construction of indicators. Additionally, surveys are based on self-reported quantities and thus are vulnerable to systematic errors that may be related to the level and characteristics of employment and income. Every effort was made to provide as coherent and unbiased a data set as possible. However, given that the database is based on

welfare and labor force surveys from around the world there still may be inconsistencies in the indicators over time due to these differences in the primary data sources that users may need to consider.

The next section outlines how surveys were selected and pre-assessed before analysis.

4. Survey selection and initial data quality checks

To ensure the quality of the quantities presented in the WWBI, all surveys proposed for inclusion underwent a four-step screening process from which a final set of surveys were selected for analysis. First, the surveys needed to have sufficient information in the key variables required for the construction of estimates (no more than 40 percent missing observations on any of the key variables used in the construction of indicators). Second, the surveys had enough observations for public sector employees to be able to construct reliable statistics. Third, the surveys were representative of employees in the entire country and not only in urban areas. Fourth, there was a clear criterion to select surveys when more than one is available for the same country and year. All four steps are briefly described below and elaborated in detail in Appendix A1:

1. To minimize potential biases emerging from estimating statistics in the presence of large swaths of missing observations, filters were designed to identify surveys with sufficient number of observations for four sets of core variables: a) employment, b) wages, c) demographics (age, gender, and rural/urban split), and d) education. Surveys with over 40 percent of observations being missing or incorrectly coded were identified. Since employment status is critical for constructing all indicators within the WWBI, surveys with more than 40 percent missing employment information (whether the individual is employed, and the nature of the employment) were excluded (666 surveys for version 2.0). For the remaining three filters, surveys were not dropped but instead only the relevant set of indicators were discarded. For example, 218 included surveys have more than 40 percent missing observations for the wage or demographic variables. For those surveys, only indicators relating to public employment were generated. This threshold has been updated from a 30 percent threshold utilized in the development of Version 1.0 based on the experience of the team and literature pointing towards 40 percent be a statistically appropriate threshold for analysis (Dong and Peng 2013; Jakobsen et al. 2017).

2. Given that the WWBI aims to capture a representative picture of public sector labor markets, ten surveys with the share of public sector employment (as a share of paid employment) lower than five percent were dropped from the database. Twenty-four samples with public sector employment observations lower than 200 were not included in the final database either.

3. In either case, the relative sample of public sector employees seems too small for the production of valid statistics. This was a decision based on the experience of the team, but small deviations from these specific thresholds do not alter our results.

4. For countries having multiple surveys within a given year, only one survey was kept for the respective year. 37 surveys were dropped after electing the more time consistent survey for the given year. As a general rule, surveys collected by the same authority or coming from the same source as other years' survey are given preference. In situations

where there is more than one survey for a given year coming from the same data source and producing similar indicators, the survey with more observations was chosen.

This process ensures that WWBI indicators generated are based on an appropriate set of surveys. However, as described in section 3, some indicators reported in the WWBI are not based on in-house processing of household surveys. Instead, these indicators, on the relative size of the wage bill and wage comparison rations reported in the database are based on external sources. Indicators on the relative size of the public sector wage bill, for example, are sources from the IMF Government Compensation and Employment Dataset, 2016. IMF (2016) which provides a detailed explanation of the construction of the indicators and the suggested caveats for using the data for cross-country analysis.

Additionally, indicators on within- and cross-country pay compression ratios in the public sector are constructed using data from the 2017 cycle of the International Comparison Program's (ICP) "Data for Researchers" dataset which is a worldwide statistical initiative to collect and compile comparable price and national accounts expenditure data and estimate purchasing power parties (PPPs) for the world's economies. The program is implemented as a global partnership of national and regional agencies and managed by the ICP Global Office at the World Bank, under the auspices of the United Nations Statistical Commission (UNSC). More information is available on the Program's website here. These indicators allow for a more focused inquiry into wage compression ratios both within and between public sectors of respective countries. Based on United Nations ISCO occupations classification system, these 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. While the latter affords a window into the inequality between the top and bottom earners within each sector and nation, the former allows for a comparison of notable occupation within the public sector of each country and across nations.

5. Identifying and dealing with outliers

While this screening ensures that the surveys selected meet a certain threshold of quality for inclusion, a second set of checks were employed on the estimates resulting from the analysis to identify outlier observations. These outliers may be the result of incorrect measurements or coding mistakes that skew sample estimates away from the population. Removing these are essential to reducing error variance and since these are expected to be distributed non-randomly, they can decrease normality. This was done using a three-step process described below with further details in Appendix A2:

1. For countries with four or more surveys between 2000 and 2020 (88 countries out of a total sample of 211), outlier observations were identified using within-country variance over year. For this set of countries, observations located more than three standard deviations away from the country's mean were marked and removed. This method identifies 833 observations (out of a total of 114,099 observations; 0.73 percent of the final dataset) as outliers from a within-country trend and removes them.

2. For countries with 3 or fewer surveys between 2000 and 2020, the above approach does not identify outliers since the standard deviation expands mechanically. Instead, observations that represented significant structural breaks from an overall country trend—where the difference between them and the country's average is greater than 20 percent

of the entire variance of the indicator—were identified and removed. This method identifies 204 outlier observations (0.18 percent of the data set).

3. Additionally, for countries with fewer than 4 observation, outliers were also identified using deviations from the mean of their region (42 new outliers identified; 0.04 percent of the data set) and income categories (52 outliers identified; 0.05 percent of the data set) – as defined by the World Bank.

Version 2.0 utilizes the above methodology for the detection and removal of outliers within the data as was used in the construction of version 1.1 and improves on the manual method utilized for the identification of outliers employed in version 1.0.

6. Categories of Variables

The WWBI 3.1 has 302 variables, grouped in three major categories.

Table 2: WWBI main categories of variables

Variable	***Source***	***Description***
Public sector and private sector employment as a share of total, paid, or formal sector employment	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Proportion of workers in each sector by using the International Classification of Status in Employment (ICSE) to define the sectors (public and private) and employment types (total, wage, and formal).*
Public sector total employment, as a share of population	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Proportion of employed population working in the public sector using the International Classification of Status in Employment (ICSE) to define the sector over total population.*
Public and private sector employment demographics: distributions by gender, age and education, industry, and location		*Proportion of workers in the public and private sectors based on key identifiers (mean and median age, male vs. females, education levels, and rural and urban divide.*
Public and private sector gender distributions of employees by occupation, industry, and wage quantile	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Gender distributions of public and private employment within major occupational groups and different wage quintiles*
Public and private sector gender distributions of employees by industry, occupation, and wage quantile	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Gender distribution in major occupational groups and different wage quintiles for both public and private sectors.*
Sample sizes of total and public sector workforces including by industry of employment	*EU-SILC, I2D2, GLD, GMD,*	*Number of total individuals employed as well as those employed in the public sector*

	LABLAC, LIS	
Public sector wage premiums	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Percentage differences in public sector wages compared to private sector wages (in local currency units) controlling for education, age, gender, and location.*
Public sector wage premiums (significance levels)	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Percentage differences in public sector wages compared to private sector wages (in local currency units) controlling for education, age, gender, and location.*
Gender pay gap	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Ratios of female to male wage (both mean and median, in local currency) for the public and private sectors.*
Pay compression ratios	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Ratios of the 90th percentile wage to the 10th percentile wage.*
Relative wages in public and private sectors across major occupations	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Ratios of wages for senior officials, professionals, and technicians (compared to clerks). It is a measure of wage progressions.*
Share of workers with benefits and social safety nets	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Share of public and private sector workers with various types of benefits – i.e. formal contracts, social security, health insurance and union membership*
Wage compression ratios in public sector, by occupation (using clerical occupations as reference)	*ICP*	*The ratio of the wages of indexed occupations to all clerical occupations (as the benchmark category).* *Occupations include: Senior officials; Judges; Government economist; Police officers; Hospital doctors; Hospital nurses; University teachers; Secondary school teachers; Primary school teacher*
Cross-country public sector wage comparison ratio, by occupation (using median and mean)	*ICP*	*The ratio of the wages of indexed occupations within the reference country to the global median (or mean) for the same category.* *Occupations include: Senior officials; Judges; Government economist; Police officers; Hospital doctors; Hospital nurses; University teachers; Secondary school teachers; Primary school teacher*

Public sector wage premiums by industry	*I2D2, GMD, GLD, LABLAC, LIS*	*Percentage differences in public sector education/healthcare industry wages compared to private sector education/healthcare wages (in local currency units) controlling for education, age, gender, and location.*
Public sector wage premiums by occupation	*I2D2, GMD, GLD, LABLAC, LIS*	*Percentage differences in public sector medical workers/teachers wages compared to private sector medical workers/teachers wages (in local currency units) controlling for education, age, gender, and location.*
Public sector wage premiums by industry (significance levels)	*I2D2, GLD, GMD, LABLAC, LIS*	*Percentage differences in public sector healthcare/education industry wages compared to private sector healthcare/education industry wages (in local currency units) controlling for education, age, gender, and location.*
Public sector wage premiums by occupation (significance levels)	*I2D2, GMD, GLD, LABLAC, LIS*	*Percentage differences in public sector medical workers/teachers wages compared to private sector medical workers/teachers wages (in local currency units) controlling for education, age, gender, and location.*
Public sector wage premium (by gender and industry)	*EU-SILC, I2D2, GMD, GLD, LABLAC, LIS*	*Wage premium for female healthcare/education employees in the public sector compared to female education/healthcare employees in the private sector.*
Gender wage premium (by industry)	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Wage premium for female healthcare/education/public admin/core public admin/public safety/social services employees in the public sector compared to male healthcare/education/public admin/core public admin/public safety/social security employees in the public sector*
Gender wage premium (by occupation)	*EU-SILC, I2D2, GLD, GMD, LABLAC, LIS*	*Wage premium for female medical workers/teachers employees in the public sector compared to male medical workers/teachers employees in the public sector*
Formality in the workforce (by industry)	*I2D2, GLD, GMD, LABLAC, LIS*	*Wage Premium for healthcare/education/public administration/core public administration/public safety/social security workforce compared to other formal employees*
Formality in the workforce (by occupation)	*I2D2, GLD, GMD, LABLAC, LIS*	*Wage Premium for medical workers/teachers compared to other formal employees*
General government wage bill as a percentage of GDP	*IMF*	*General government wage bill in proportion to country GDP (based on PPP; 2009 dollars).*

General government wage bill as a percentage of Government expenditures	*IMF*	*General government wage bill in proportion to total general government expenditures (based on PPP; 2009 dollars).*

7. Detailed Variable Description

7.1 *Basic identification variables/classifications*

Table 3: WWBI Basic identification variables

Variable Name	*Source*	*Description*
Country name	*World Bank*	*World Bank classification nomenclature (described here) with concordance to other widely-accepted systems (often referred to as the "ISO 3166-1" system) such as that of the IMF (here) and OECD (here). Data cleaning and editing were conducted to ensure country equivalence, especially in unique cases (e.g. "Gambia" vs. "The Gambia," "Burma" vs. "Myanmar," etc.). Income classifications were also updated using the World Bank's schema (viewable here). Other income and regional classifications were not used.*
Country code		
Income Group		
Region		

7.2 The classification of labor and employment status

The construction of all indicators included in the WWBI rests on the precise identification of employed individuals, paid employees, and public paid employees. These three definitions essential for constructing indicators on the (absolute and relative) size of the public sector workforce and lay the foundation of all disaggregated indicators and the wage analysis. These are based on the I2D2 dataset as defined below:

Table 4: WWBI Labor definitions

Variable name and values	*Variable Description*
Labor status • *Employed* • *Unemployed* • *Non-in-labor force*	*Variable is constructed for all persons administered the labor module in each survey. In the original household surveys, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. When constructing the WWBI, we use the international standard of 15 as the lower age cutoff.[3] Therefore, we only include respondent aged 15 and above in the analysis. All persons are considered active in the labor force if they presently have a job (formal or informal, i.e. are employed) or do not have a job but are actively seeking work (i.e. unemployed). Employment and unemployment definitions are taken from the survey itself.*
Employment status • *Paid Employee*	*Variable is constructed for those who are working (labor status=employed).*

[3] Given the heterogeneous application of retirement ages across countries, no upper age cutoff was utilized. This was done in order to ensure that the analysis in order to maintain the full roster of public sector professionals, who are, on average, older than their private sector counterparts.

• *Non-Paid Employee* • *Employer* • *Self-employed* • *Other, workers not classifiable by status*	*Definitions taken from the International Labor Organization's Classification of Status in Employment with some revisions to take into account the data available. Classifies the main job employment status of any individual with a job (labor status=employed).* ***Paid employee** includes anyone whose basic remuneration is not directly dependent on the revenue of the unit they work for, typically remunerated by wages and salaries but may be paid for piece work or in-kind. The 'continuous' criteria used in the ILO definition is not used here as data are often absent and due to country specificity.* ***Non-paid employee** includes contributing family workers are those workers who hold a self-employment job in a market-oriented establishment operated by a related person living in the same households who cannot be regarded as a partner because of their degree of commitment to the operation of the establishment, in terms of working time or other factors, is not at a level comparable to that of the head of the establishment.* ***Employer** is a business owner (whether alone or in partnership) with employees. If the only people working in the business are the owner and contributing family workers, the person is not considered an employer (as has no employees) and is, instead classified as own account.* ***Own account or self-employment** includes jobs are those where remuneration is directly dependent from the goods and service produced (where home consumption is considered to be part of the profits) and have not engaged any permanent employees to work for them on a continuous basis during the reference period.* ***Members of producers' cooperatives** are workers who hold a self-employment job in a cooperative producing goods and services in which each member takes part on an equal footing with other members in determining the organization of production, sales and/or other work of the establishment, the investments and the distribution of the proceeds of the establishment amongst the members.* ***Other, workers not classifiable by status** include those for whom insufficient relevant information is available and/or who cannot be included in any of the preceding categories.*
Sector of activity • *Public sector, Central Government, Army, NGO, State owned company* • *Private*	*Variable is constructed for all persons administered this module in each questionnaire. Classifies the main job's sector of activity of any individual with a job (labor status=employed) and is missing otherwise.* ***Public sector** includes central government, non-governmental organizations, armed forces, and state-owned company.* ***Private sector** is that part of the economy which is both run for private profit and is not controlled by the state.* *Information is mainly missing for people below working age, unemployed and for people out of the labor force. Other missing values are allowed.*
Wage	*Wage from main job (job to which the person dedicated most time in the week preceding the survey). This excludes tips, bonuses, and other payments.*

	For all those with self-employment or owners of own businesses, this should be net revenues (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes. *By definition non-paid employees should have zero wage.* *For some surveys wage for self-employment or for owners of own business is missing.*

7.3 Employment variables by sector

Table 5: WWBI Public employment variables

Variable Name	***Description***
Public sector employment as a share of total employment	*Number of public sector paid employees/Number of all employed individuals (including: paid employee, non-paid employee, employer, self-employment, and other workers)*
Public sector employment as a share of total employment by gender: male/ female	*Number of female (male) public sector paid employees/Number of employed females (male)*
Public sector employment as a share of total employment by location: rural/ urban	*Number of public sector paid employees in rural (urban) area /Number of all employed individuals in rural (urban) area*
Public sector total employment as a share of population	*Number of public sector workers/Total population in the country*
Public sector employment as a share paid employment	*Number of public sector paid employees/ Number of all paid employees (i.e. all employees in wage employment)*
Public sector employment as a share of paid employment by gender: male/ female	*Number of female (male) public sector paid employees/ Number of female (male) paid employees*
Public sector employment as a share of paid employment by location: rural/ urban	*Number of public sector paid employees in rural (urban) area / Number of all paid employees in rural (urban) area*
Public sector employment as a share of formal employment	*Number of public sector paid employees/ Number of formal employees. Formal employment is defined by having access to at least one of following benefits (contract, health insurance, union membership, or social security)*

7.4 Worker demographic breakdowns (non-gender-related)

Table 6: WWBI Demographic variables

Variable Name	***Description***
Mean (median) age of public (private) paid employees	*The mean and median age of public and private paid employees are calculated to provide an age profile for public and private sector employees.*
Individuals with tertiary education (no education, primary education, secondary education) as a share of public (private) paid employees	*Number of public (private) paid employees with tertiary education (no education, primary education, secondary education / Total number of public (private) paid employees*
Proportion of total employees with tertiary education working in the public sector	*Number of tertiary education holders work as a public paid employees/Number of employed individuals with tertiary education*
Rural residents as a share of public (private) paid employees	*Number of public (private)paid employees live in rural area/ Total number of public (private) paid employees*

The demographic information used to construct the above indicators are derived from the I2D2 survey, the descriptions are as follows:

Table 7: I2D2 demographic categories

Variable name and values	***Variable Description***
Gender • *Male* • *Female*	*Self-identification of gender. Note that this is based on in-county survey standards and thus may not include options other than "Male" or "Female."*
Age • *15-99 (continuous)*	*Age is reported in years and truncates to exclude individuals (children) under 15 years of age and thus outside the employable population (according to international standards).*
Education • *No education* • *Primary* • *Secondary* • *Tertiary*	*This variable follows the simpler convention than compared to the "years of education" since the latter is country specific as not all countries require the same number of school years to complete a given level. Therefore, the data is collapsed to four main categories based on the highest level of education completed, level and/or years of education, depending on the nature of the questionnaire.* ***No education*** *includes people with no education listed or primary education listed as incomplete. If the variable is based on years of schooling, this is determined only after a review of country specific literature.* ***Primary*** *completed implies that one completed the stipulated primary education by undertaking an exam or test, where this exists. Otherwise, refers to having completed the highest grade in this level of education.* ***Secondary*** *completed implies that one completed the stipulated secondary (middle-school) education by undertaking an exam or*

	test, where this exists. Otherwise, refers to having completed the highest grade in this level of education. ***Tertiary*** *education includes any identification of post-secondary, college, or university education successfully completed after secondary level of education* *Religious or vocational schools designed primarily to impart religious education with some basic literacy, or technical or vocational training are treated on a case basis based on within-country designation.*
Urban/ Rural • *Urban* • *Rural*	*Country-specific as each country defines this jurisdiction according to its own criterion. In transition economies where 'semi-urban' is a recognized category which includes 'villages of the town type' this has been collapsed into the 'urban' category. Further information on local administrative units can be found here.*

7.5 Worker demographics (gender-based)

In addition to basic demographic trends, the WWBI is especially interested in the role of women in the public and private sectors. Thus, we have gathered several additional variables to track gender dynamics between sectors and across countries. WWBI includes distributions of women in different occupational roles in both public and private sectors – e.g. Senior officials, Technicians, Clerks, etc. Percentage of females is also reported for different wage quintile groups in both public and private sectors.

Table 8: WWBI Gender-based demographic variables

Variable Name	***Description***
Females as a share of public paid employees by occupation: senior officials (professionals, technicians, clerks, Elementary occupations workers) in public (private) sector	*Number of female public paid employees who work as senior official (professionals, technicians, clerks, Elementary occupations workers) / Number of public paid employees work as senior official (professionals, technicians, clerks, Elementary occupations workers)*
Females as a share of public paid employees by wage quintile: 1st - 5th	*Number of female public paid employees in 1st wage quintile (2nd, 3rd, 4th, 5th wage quintile) / Number of public paid employees in 1st wage quintile (2nd, 3rd, 4th, 5th wage quintile)*
Females as a share of public paid employees	*Number of female public (private) paid employees/ Total number of public (private) paid employees* *The indicator here is based on paid employee only. Thus, we exclude working individual with other employment type (self-employed, non-paid employee, employer, etc.) when calculating this indicator. It should not be viewed as the share of female in labor force.*

The I2D2 provides data on employment by occupation for the International Standard Classification of Occupations (ISCO) major occupational categories (e.g. senior officials, professionals,

technicians etc.). We report data on a subset of the occupations that can be compared between the public and private sectors. Additionally, we report the relative size of these indicators considering the whole public sector workforce and the formal workforce along with the proportion of female workers occupied in these major categories (Table 9B).

The detailed list of occupation classification and additional indicators featuring this classification are listed below:

Table 9A: Occupation classifications

Numeric classifications	*Variable description*
Occupation • *Managers* • *Professionals* • *Technicians and associate professionals* • *Clerical support workers* • *Elementary occupations*	*Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and upper age cutoff) at which information is collected will vary from country to country. The variable classifies the main job of any individual with a job (labor status=employed) and is missing otherwise. As most surveys collected detailed information and then coded it, and the original data is not in the databases, no attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO) 88. In the case of different classifications re-coding has been done to best match the ISCO-88.* ***Managers** plan, direct, coordinate and evaluate the overall activities of enterprises, governments, and other organizations. This classification includes chief executives, senior officials, legislators, and managers of any kind.* ***Professionals** increase the existing stock of knowledge; apply scientific or artistic concepts and theories; teach about the foregoing in a systematic manner; or engage in any combination of these activities. This classification includes health, teaching, business professionals etc.* ***Technicians and associate professionals** perform technical and related tasks connected with research and the application of scientific or artistic concepts and operational methods, and government or business regulations.* ***Clerical support workers** record, organize, store, compute and retrieve information, and perform clerical duties in connection with money-handling operations, travel arrangements, requests for information, and appointments. This classification includes general and keyboard clerks, customer services, other clerical support workers etc.* ***Elementary occupations** involve the performance of simple and routine tasks which may require the use of hand-held tools and considerable physical effort. This classification includes cleaners, agricultural labor, labors in mining, construction, manufacturing, and transport etc.*

Table 9B: Relative size and proportion of female workers in major occupations

Variable name	Variable description
Public sector "manager/ professional/technician/clerk, elementary worker", as a share of formal employment	*Number of public employees who work as professionals, technicians, clerks, Elementary occupations workers / Number of formal employees. Formal employment is defined by having access to at least one of following benefits (contract, health insurance, union membership, or social security)*
Public sector "manager/ professional/technician/clerk, elementary worker", as a share of paid employment	*Number of public employees who work as professionals, technicians, clerks, Elementary occupations workers / Number of public paid employees*
Public sector "manager/ professional/technician/clerk, elementary worker", as a share of total employment	*Number of public employees who work as professionals, technicians, clerks, Elementary occupations workers / Number of all employed individuals (including: paid employee, non-paid employee, employer, self-employment, and other workers)*
Public sector "manager/ professional/technician/clerk, elementary worker", as a share of paid employment, by gender: female	*Number of female "manager/ professional/technician/clerk, elementary worker"/ Number of public paid employees*

7.6 Broad Industry variables (healthcare, education, public administration, core public administration)

WWBI also reports data on six industries – healthcare, education, public administration, core public administration, public safety and social services- that can be compared between the public and private sectors (expect for those in public administration and subindustries).

We used a specific question on the industry of employment to identify employees in the six industries. This is the case for data sourced from the GLD, GMD, I2D2, LABLAC, and LIS micro-data repositories. As in the occupational classification, the only exception to this is the data sourced from Eurostat's European Union Statistics on Income and Living Conditions (EU-SILC) for countries within the European Economic Area (EEA). For these samples, NACE Rev. 2 industry classifications "O" which covers public administration, defense and compulsory social security, P which accounts for human health services and "Q" covering education services. Although the aggregation of these three sectors leaves out the private sector employment within these three industries, it provides a fair approximation to the government/public sector, especially for countries in the European Union and is standard practice within the literature (Christofides and Michael 2013; de Castro, Salto, and Steiner 2013; European Commission 2014; Giordano et al. 2015).

For the majority of samples sourced from the I2D2 and LIS micro-data repositories, the classification is based on The International Standard Industrial Classification of All Economic

Activities (ISIC). The ISIC is the international classification of productive activities and provides a comprehensive framework for economic data collection and reporting. These economic activities are subdivided in a four-level hierarchical structure. All categories are mutually exclusive.

Ideally, countries can provide data according to ISIC at all levels of the classification for the purposes of international comparability. However, the quality and comparability of the statistics largely depends on the correctness of the codes, the type of statistical source, the degree of elaboration that countries find it necessary and/or feasible to classify their economic activity data, and confidentiality considerations (UNSD 2007). Therefore, in our analysis we construct broad classification categories of healthcare, education, public administration, core public administration, public safety and social services industries in order to avoid misclassifications.

At the highest level of ISIC, some sections can be easily compared to the previous ISIC revisions. ISIC correspondence tables are provided below:

Table 10: ISIC industry classifications (healthcare, education, public administration ,core public administration, public safety and social security)

Numeric classifications	*Variable description*
ISIC 2.0	***910 Public administration and defense*** ***931 Education activities*** ***933 Medical, dental, and other health services***
ISIC 3.1	***75 Public administration and defense; compulsory social security*** ***Note:*** *The division includes general administration (e.g., executive, legislative, financial administration etc. at all levels of government) and supervision in the field of social and economic life - defense, justice, police, foreign affairs etc. - management of compulsory social security. Activities classified elsewhere in ISIC do not fall under division 75 even if carried out by public administrations. For example, administration of the school system (regulations, checks, curricula) falls under division 75, but teaching itself does not.* ***751 Administration of the State and the economic and social policy of the community*** ***Note****: The division includes: executive and legislative administration of central, regional and local bodies; administration and supervision of fiscal affairs (operation of taxation schemes; duty/tax collection on goods and tax violation investigation; customs administration; budget implementation and management of public funds and public debt: raising and receiving of moneys and control of their disbursement; administration of overall (civil) R&D policy and associated funds; administration and operation of overall economic and social planning and statistical services at the various levels of government. This division excludes administration of R&D policies intended to increase personal well-being and of associated funds; administration of R&D policies intended to improve economic*

	performance and competitiveness and administration of defense-related R&D policies and of associated funds. ***752 Provision of services to the community as a whole*** ***Note:*** *The division includes: Foreign affairs, defense activities, public order and safety activities.* ***753 Compulsory social security activities*** ***80 Education*** ***Note****: The division includes public as well as private education at any level or for any profession, oral or written as well as by radio and television or other means of communication. It includes education by the different institutions in the regular school system at its different levels as well as adult education, literacy programmes and others. Also included are military schools and academies, prison schools etc. at their respective levels.* *For each level of initial education, the classes include special education for physically or mentally handicapped pupils. The breakdown of the categories in this division is based on the level of education offered as defined by the levels of ISCED 1997. This division excludes education primarily concerned with recreational activities such as bridge or golf* ***85 Health and social work*** ***Note****: This division includes the provision of health care by diagnosis and treatment and the provision of residential care for medical and social reasons, as well as the provision of social assistance, such as counselling, welfare, child protection, community housing and food services, vocational rehabilitation and childcare to those requiring such assistance. Also included is the provision of veterinary services.*
ISIC 4.0	***84 Public administration and defense; compulsory social security*** ***Note:*** *This division includes activities of a governmental nature that are normally carried out by the public administration, including the enactment and judicial interpretation of laws and their pursuant regulation; the administration of programmes based on them; legislative activities; taxation; national defense; public order and safety; immigration services; foreign affairs; and the administration of government programs* **841 Administration of the State and the economic and social policy of the community** **Note:** *This division includes general public administration activities; regulation of the activities of providing health care, education, cultural services and other social services, excluding social security and regulation of and contribution to more efficient operation of businesses.* ***842 Provision of services to the community as a whole*** *Note: This group includes foreign affairs, defense and public order and safety activities. Foreign affairs. This class includes: administration and operation of the ministry of foreign affairs and*

	diplomatic and consular missions stationed abroad or at offices of international organizations administration, operation and support for information and cultural services intended for distribution beyond national boundaries aid to foreign countries, whether or not routed through international organizations provision of military aid to foreign countries management of foreign trade, international financial and foreign technical affairs This class excludes: international disaster or conflict refugee services. Defense activities This class includes: administration, supervision and operation of military defense affairs and land, sea, air and space defense forces such as: combat forces of army, navy and air force, engineering, transport, communications, intelligence, material, personnel and other non-combat forces and commands, reserve and auxiliary forces of the defense establishment, military logistics (provision of equipment, structures, supplies etc.), health activities for military personnel in the field administration, operation and support of civil defense forces support for the working out of contingency plans and the carrying out of exercises in which civilian institutions and populations are involved administration of defense-related R&D policies and related funds. Public order and safety activities. This class includes: administration and operation of regular and auxiliary police forces supported by public authorities and of port, border, coastguards and other special police forces, including traffic regulation, alien registration, maintenance of arrest records firefighting and fire prevention: administration and operation of regular and auxiliary fire brigades in fire prevention, firefighting, rescue of persons and animals, assistance in civic disasters, floods, road accidents etc. administration and operation of administrative civil and criminal law courts, military tribunals and the judicial system, including legal representation and advice on behalf of the government or when provided by the government in cash or services. ***843 Compulsory social security activities*** ***Note:*** *This class includes funding and administration of government-provided social security programs: sickness, work-accident and unemployment insurance, retirement pensions, programs covering losses of income due to maternity, temporary disablement, widowhood etc.* ***85 Education*** ***Note:*** *This division includes education by the different institutions in the regular school system at its different levels as well as adult education, literacy programmes etc. Also included are military schools and academies, prison schools etc. at their respective levels. The section includes public as well as private education. For each level of initial education, the classes include special education for physically or mentally handicapped pupils. The breakdown of the categories in this section is based on the level of education offered as defined by the levels of ISCED 1997. This section also includes instruction primarily concerned with sport and recreational activities such as bridge or golf and education support activities.*

	86–88 Human health and social work activities ***Note:*** *This division includes the provision of health and social work activities. Activities include a wide range of activities, starting from health care provided by trained medical professionals in hospitals and other facilities, over residential care activities that still involve a degree of health care activities to social work activities without any involvement of health care professionals.* *This division also includes the provision of residential care combined with either nursing, supervisory or other types of care as required by the residents. Facilities are a significant part of the production process and the care provided is a mix of health and social services with the health services being largely some level of nursing services.* *This division also includes the provision of a variety of social assistance services directly to clients. The activities in this division do not include accommodation services, except on a temporary basis.*

We paid great attention to identifying the coding structure of each sample in accordance with different ISIC revisions. One of the strategies we have used was identifying additional division codes or gaps among ISIC 2.0, ISIC 3.1, and ISIC 4.0 revisions. Some gaps in revisions have been provided to allow countries to introduce division-level categories without requiring a complete change of the ISIC coding. For example, in ISIC 4.0, the following division code numbers have been left unused: 04, 34, 40, 44, 48, 54, 57, 67, 76, 83 and 89. We have checked for these number sequences in the coding structure to identify samples coded with ISIC 4.0.

For some Mexican samples, we have coded the industry variables based on The North American Industry Classification System (NAICS) system. It was adopted in 1997 to replace the Standard Industrial Classification (SIC) system and uses 6-digit code. The following codes have used to identify relevant industries for Mexico samples: 61- Educational services, 62- Healthcare and social assistance, and 91/92- Public administration.

Some samples were found to use ISIC Rev 3.1 or ISIC Rev 4.0 Sections as opposed to Divisions.

Table 11: Industry focused public employment variables

Variable Name	***Description***
Public sector employment as a share of paid employment by industry	*Number of individuals employed in the public sector / all paid employees within (education/healthcare/public administration/core public administration/public safety/social security) industry.*
Public sector employment as a share of total employment by industry	*Number of individuals employed in the public sector / total employees (*Paid Employee; Non-Paid Employee; Employer; Self-employed; Other, workers not classifiable by status) *within (education/healthcare/public administration/core public administration/public safety/social security) industry.*

Public sector employment as a share of formal employment by industry	*Number of individuals employed in the public sector / all formal employees (with social security, insurance, and contract) within (education/healthcare/public administration/core public administration/public safety/social security) industry.*

Table 12: Industry focused demographic variables (non-gender)

Variable Name	***Description***
***Mean age of public paid employees* by industry**	*The mean age of public paid employees in healthcare, education, and public administration industries are calculated to provide an age profile for public sector employees within each industry.*
***Mean age of private employees* by industry**	*The mean age of private paid employees in healthcare, and education industries are calculated to provide an age profile for public sector employees for both industries.*
Employees *with tertiary education as a share of* all *public sector* employees by industry.	*Number of public healthcare/education/public administration/core public administration/public safety/social security paid employees with tertiary education / Total number of public healthcare/education/public administration/core public administration/public safety/social security paid employees*
Employees in rural areas as a share of all public sector *employees* by industry.	*Number of public healthcare/education/public administration/core public administration paid employees living in rural areas/ Total number of public healthcare/education/public administration/core public administration/public safety/social security paid employees*

In addition to basic demographic trends of healthcare, education, and public administration employees, the WWBI is also interested in the role of women within these industries and how it compares across public and private sectors. Thus, we have gathered several additional variables to track gender dynamics between sectors and industries across countries.

Table 13: Industry focused gender-based variables

Variable Name	***Description***
Female healthcare/education paid employees in the public sector as a share of all	*Number of female healthcare/educations paid employees in the public sector/ Number of all female healthcare/education industries paid employees*

healthcare/education paid employees	

7.7 Key occupation variables (medical workers, teachers)

In addition to the six industries, the WWBI also reports data on key occupational variables- medical workers (excluding non-medical healthcare employees) and teachers (excluding non-teaching education workers)- that can also be disaggregated into public and private sectors and gender. We chose these three occupations because the vast majority of public sector workers across most of the countries in our database are employed in them, according to our own analysis.

To construct these variables, we use the occupation in employment question to identify the medical workers and teachers in each survey coming from GLD, GMD, I2D2, LABLAC and LIS harmonization efforts. All these surveys use the International Standard Classification of Occupations (ISCO) to classify productive activities. This classification organizes jobs into a clearly defined set of groups according to the tasks and duties undertaken in the main activity, based on International Labor Organization standards. As previously mentioned, the construction of the occupation variables was not possible for the data sourced from Eurostat's European Union Statistics on Income and Living Conditions (EU-SILC) for countries within the European Economic Area (EEA), which follows a different classification.

Most of the surveys in our database use either ISCO-88 or ISCO-08 classification structures. Since the launch of ISCO-08, many countries have been updating their national classification based on the old structure to improve alignment with the new international statistical standard. The updating between these two versions did not change the basic principles and top structure of the categories but significant structural changes were made in some specific areas. Still, most sections can still be easily compared between the two revisions (details in Table 14).

Table 14: ISCO occupation classifications (medical workers and teachers)

Numeric classifications	***Variable description***
ISCO-88	**22 Life science and health professionals** ***Note:*** *This division includes life science professionals, health professionals and nursing and midwifery professionals.* ***23 Teaching professionals*** ***Note****: The division includes college, university and higher education teaching professionals; secondary education teaching professionals; primary education teaching professionals; pre-primary education teaching professionals and special education teaching professionals*.

ISCO-08	***22 Health Professionals*** ***Note:*** *This division includes medical doctors, nursing and midwifery professionals, paramedical practitioners, other health professionals.* ***23 Teaching professionals*** ***Note:*** *This division includes University and higher education teachers; Vocational Education teachers; Secondary education teachers; Primary school and Early Childhood teachers and Other Teaching professionals.*

As we did with the industrial classification variables, we again paid great attention to identifying the coding structure of each sample in accordance with the two different ISCO revisions. With this classification of occupations, we were able to create major occupation focused public employment variables, to understand the relative size of each occupation in terms of total employment, paid employment, and formal employment (see Table 15) and in order to characterize these workers in terms of educational background and location (Table 16) and understand the role of women workers in these occupations both for public and private sectors (Table 17).

Table 15: Occupation focused public employment variables

Variable Name	***Description***
Public sector employment as a share of paid employment by occupation	*Number of individuals employed in the public sector / all paid medical workers/teachers*
Public sector employment as a share of total employment by occupation	*Number of individuals employed in the public sector / total employees (*Paid Employee; Non-Paid Employee; Employer; Self-employed; Other, workers not classifiable by status) *within medical workers/teachers occupations*
Public sector employment as a share of formal employment by occupation	*Number of individuals employed in the public sector / all formal employees (with social security, insurance, and contract) within medical workers/teachers occupations.*

Table 16: Occupation focused demographic variables (non-gender)

Variable Name	***Description***
Mean age of public paid employees **by occupation**	*The mean age of public paid "medical workers/teachers" employees are calculated to provide an age profile for public sector employees within each occupation.*
Mean age of private employees **by occupation**	*The mean age of private paid "medical workers/teachers" are calculated to provide an age profile for public sector employees for the three occupations.*
Employees *with tertiary education as a share of* all	*Number of public paid "medical workers/teachers" with tertiary education / Total number of paid "medical workers/teachers"*

***public sector* employees by occupation.**	
Employees in rural areas as a share of all public sector *employees* by occupation.	*Number of public "medical workers/teachers" living in rural areas/ Total number of public "medical workers/teachers"*

Table 17: Occupation focused gender-based variables

Variable Name	***Description***
***Female paid "medical workers/teachers" in the public sector as a share of all paid** "medical workers/teachers"*	*Number of female paid "medical workers/teachers" in the public sector/ Number of all female paid "medical workers/teachers"*

7.8 Wage differentials

As noted above in section 2,3, wage information in the surveys are reported in each country's local currencies (Local Currency Units, LCU) and a diverse array of periodicity. Great care is taken to identify the exact frequency of income for each individual within the surveys and converting all wages to weekly wage based on well-established conversion factors as described below:

If wage frequency is Hourly, *Monthly Wage = Raw Wage * (hours worked/week * 4.345)*
If wage frequency is Daily, *Monthly Wage = Raw Wage * (5 * 4.345)*
If wage frequency is Weekly, *Monthly Wage = Raw Wage * 4.345*
If wage frequency is Biweekly, *Monthly Wage = Raw Wage * (4.345 / 2)*
If wage frequency is Bimonthly, *Monthly Wage = Raw Wage * 0.5*
If wage frequency is Monthly, *Monthly Wage = Raw Wage*
If wage frequency is Quarterly, *Monthly Wage = Raw Wage * 1/3*
If wage frequency is Biannually, *Monthly Wage = Raw Wage * 1/6*
If wage frequency is Annually, *Monthly Wage = Raw Wage * 1/12*

After the initial harmonization of all wage data from the surveys the following wage variable transformations are made for analysis within the WWBI:

wpm (wage per month) = Monthly Wage
wpw (wage per week) = Monthly Wage / 4.345

To control for the effect of possibly spurious outliers, the wage variables are winsorized by limiting extreme values in the survey data at the 0.01 percent level.

In order to make cross-country comparisons, we construct pay compression ratio and gender wage differentials. WWBI also reports these gender wage differentials across different education level within the public and private sectors.

Table 18: WWBI Wage differential variables

Variable Name	***Source***	***Description***
Pay compression ratio in public (private) sector	*I2D2*	*90th percentile of weekly wage for public (private) paid employees/10th percentile of weekly wage for public (private) paid employees*
Female to male wage ratio in public (private) sector using mean (median)	*I2D2*	*Mean (Median) weekly wage of female public (private) paid employees / Mean (Median) weekly wage of male public (private) paid employees*

7.9 Wage premium

Estimating public sector wage competitiveness compared to the private sector is methodologically complicated. The standard approach in the academic literature is to measure differences in total compensation between the public and private sectors for statistically similar workers in similar jobs. Given the demographic differences of workers between the two sectors, this approach ideally requires controlling for observable worker characteristics, such as age, education, work experience, and gender that impact human capital and therefore earnings; accounting for unobserved characteristics such as ability, risk aversion, and public service motivation; and controlling for occupations given that the similar workers can have very different responsibilities in different occupations. A simple raw comparison of average wages in the private and in the public sector is misleading as public sector workers are older and more educated than their private sector counterparts, have different career objectives and motivations, and also work on occupations that are not well represented in the private sector.

In order to estimate public sector wage premium, Mincerian earnings regressions were utilized specified with a dummy variable indicating the sector of the individual.[4] The basic specification is as follow:

$$\boldsymbol{logw_i = \alpha + \beta \cdot PUBLIC_i + X_i \cdot \gamma + \epsilon_i} \qquad \textbf{(1)}$$

Where β is the adjusted public-private wage differential; $logw_i$ is logged weekly wages in local currency of employee i; $PUBLIC_i$ is a dummy=1 if wage employee works in public sector; and X_i is a vector of standard controls consisting of age, age squared, level of education, location (urban/rural), and gender.

[4] The two main empirical approaches in the literature are the Mincerian wage regression with a dummy variable indicating whether the worker is employed in the public sector or private sector; and the Oaxaca-Blinder decomposition which does not assume that the returns to education, gender, age and other observable worker characteristics are the same in the public and private sector. The latter method decomposes the wage differential into a part that can be explained as resulting from worker endowments, and an unexplained part presumably due to economic rents that the public sector enjoys. The two approaches in general give similar results (Gittleman and Pierce 2011), so we use the dummy variable method as it is simpler to present. To allow the public sector earnings differential to vary between individuals we estimate Mincer-style wage gaps by gender, age, occupation, skill level and other characteristics.

Reported premiums are transformed based on equation (2) as the untransformed $\hat{\beta}$ only gives an approximation of the actual premium and the discrepancy becomes larger when the $\hat{\beta}$ > ± 20%. Within the WWBI, 7497 observations for wage premia are reported across 13 indicators and all countries and years. This simple transformation allows for a more precise estimation of premia.[5]

$$\%\Delta \boldsymbol{y} = \mathbf{100} * (\boldsymbol{e}^{\boldsymbol{\beta}_1 \Delta \boldsymbol{x}} - \mathbf{1}) \qquad \textbf{(2)}$$

The decision to use a simple regression specification was in part due to a relative tradeoff between a well-specified equation and an inability to provide a large set of observations for multiple countries using a consistently defined specification across countries. For a similar reason, there is no variable in the raw data that may reasonably allow for a more precise instrument for wage differentials while controlling for selection or endogeneity. Additionally, incomes within the data are also limited to self-reported wages, and do not include bonuses, allowances, and in-kind payments, which can be significant in the public sector. Certain surveys do include information on work benefits, such as health insurance and social security, but these are not monetized and cannot be added to wages to provide an estimate of total compensation.

Table 19: WWBI Wage premium variables

Variable Name	***Description***
Public sector wage premium (compared to all private paid employees)	*We estimate the equation separately for each country using OLS with robust standard errors. The data used includes only salaried employees with a positive wage. We use additional specifications that restrict the private sector to formal wage employees only and other specifications that interact $PUBLIC_i$ with gender, education and occupation to estimate public wage premium by gender, education and occupation.*
Public sector wage premium (compared to formal wage employees)	
Public sector wage premium by gender	
Public sector wage premium (compared to formal wage employee) by occupation	
Public sector wage premium (compared to formal wage employee) by education	

7.10 Industry specific wage premiums

In order to estimate wage premium by industry, we use Mincerian earnings regressions with a dummy variable indicating whether the worker is employed within that industry in the public sector or private sector. Our specification is:

$$\boldsymbol{logw_i} = \boldsymbol{\alpha} + \boldsymbol{\beta} \cdot \boldsymbol{IndustryPublic_i} + \boldsymbol{X_i} \cdot \boldsymbol{\gamma} + \boldsymbol{\epsilon_i} \qquad \textbf{(1)}$$

[5] The algebraic expression of the transformation of these premia are documented in Appendix A3

Where β is the adjusted public-private wage differential by industry; $logw_i$ is logged weekly wages in local currency of employee i; $IndustryPublic_i$ is a dummy=1 if wage employee works within that industry in public sector; and X_i is a vector of standard controls consisting of age, age squared, level of education, location (urban/rural), and gender.

Reported premiums are transformed based on equation (2) as the untransformed $\hat{\beta}$ only gives an approximation of the actual premium and the discrepancy becomes larger when the $\hat{\beta}$ > ± 20%.

$$\%\Delta \boldsymbol{y} = \mathbf{100} * (\boldsymbol{e}^{\boldsymbol{\beta}_1 \Delta \boldsymbol{x}} - \mathbf{1}) \qquad \textbf{(2)}$$

7.11 Occupation specific wage premiums

To estimate wage premiums by occupation, we also use Mincerian earnings regressions with a dummy variable indicating whether the worker is employed as medical worker or teacher in the public sector or private sector. We also take into consideration the sample sizes of these indicators to be able to keep some level of precision and statistical significance in the wage premiums estimates. In that regard, we set a threshold of 20 observations per occupation indicator before running the models. Our specification is:

$$\boldsymbol{logw_i = \alpha + \beta \cdot OccupationPublic_i + X_i \cdot \gamma + \epsilon_i} \qquad \textbf{(1)}$$

Where β is the adjusted public-private wage differential by occupation (medical worker, teacher); $logw_i$ is logged weekly wages in local currency of employee i; $OccupationPublic_i$ is a dummy=1 if wage employee works as medical worker or teacher in public sector; and X_i is a vector of standard controls consisting of age, age squared, level of education, location (urban/rural), and gender.

Similar to the industry wage premiums, reported premiums are transformed based on equation (2) as the untransformed $\hat{\beta}$ only gives an approximation of the actual premium and the discrepancy becomes larger when the $\hat{\beta}$ > ± 20%.

$$\%\Delta \boldsymbol{y} = \mathbf{100} * (\boldsymbol{e}^{\boldsymbol{\beta}_1 \Delta \boldsymbol{x}} - \mathbf{1}) \qquad \textbf{(2)}$$

7.12 Gender wage gap by industry

In order to estimate the gender wage gap, we run Mincerian earnings regressions, similar to the above, except specified with a dummy variable indicating the gender of the individual. The basic specification is as follow:

$$\boldsymbol{logw_i = \alpha + \beta \cdot FemIndPub_i + X_i \cdot \gamma + \epsilon_i} \qquad \textbf{(1)}$$

Where β is the adjusted gender wage differential by industry within the public sector; $logw_i$ is logged weekly wages in local currency of employee i; $FemIndPub_i$is a dummy=1 if female employee works within that industry in public sector; and X_i is a vector of standard controls consisting of age, age squared, level of education, and location (urban/rural).

Reported premiums are transformed based on equation (2) as the untransformed $\hat{\beta}$ only gives an approximation of the actual premium and the discrepancy becomes larger when the $\hat{\beta}$ > ± 20%.

$$\%\Delta \boldsymbol{y} = \mathbf{100} * (\boldsymbol{e}^{\boldsymbol{\beta}_1 \Delta \boldsymbol{x}} - \mathbf{1}) \quad \textbf{(2)}$$

7.13 Gender wage gap by occupation

Similar to the above, we also model the gender wage gap by occupation running Mincerian earning regressions including a dummy variable indicating the gender of the individual. As explained above, in order to keep some level of precision and statistical significance, we only consider surveys with at least 20 observations in each occupational indicator to estimate the wage gaps. The basic specification is as follow:

$$\boldsymbol{logw_i = \alpha + \beta \cdot FemOccupPub_i + X_i \cdot \gamma + \epsilon_i} \quad \textbf{(1)}$$

Where β is the adjusted gender wage differential by occupation within the public sector; $logw_i$ is logged weekly wages in local currency of employee i; $FemOccupPub_i$is a dummy=1 if female employee works as a doctor, nurse or teacher in public sector; and X_i is a vector of standard controls consisting of age, age squared, level of education, and location (urban/rural).

Reported premiums are again transformed based on equation (2) as the untransformed $\hat{\beta}$ only gives an approximation of the actual premium and the discrepancy becomes larger when the $\hat{\beta}$ > ± 20%.

$$\%\Delta \boldsymbol{y} = \mathbf{100} * (\boldsymbol{e}^{\boldsymbol{\beta}_1 \Delta \boldsymbol{x}} - \mathbf{1}) \quad \textbf{(2)}$$

For both industry and occupation indicators, we estimate above equations separately considering the industry and occupation of the workers. Detailed variable definitions are shown in Table 20 and Table 21.

Table 20: Industry-focused wage premium variables

Variable Name	***Description***
Public sector wage premium (compared to all private paid employees) by industry	*We estimate the equation separately by each industry (healthcare/education/public administration/core public administration/public safety/social security) for each country using OLS with robust standard errors.* *The data used includes only salaried employees with a positive wage.*
Public sector wage premium (compared to formal wage employees) by industry	
Public sector wage premium by gender	
Gender wage premium by industry	
Public sector wage premium (compared to formal wage employee) by education	

Table 21: Occupation-focused wage premium variables

Variable Name	***Description***

Public sector wage premium (compared to all private paid employees) by occupation	*We estimate the equation separately by each occupation (medical worker/teacher) for each country using OLS with robust standard errors.* *The data used includes only salaried employees with a positive wage.*
Public sector wage premium (compared to formal wage employees) by occupation	
Gender wage premium by occupation	

7.14 Benefits

Table 22: WWBI Benefits variables

Variable Name	*Description*
Share of public (private) paid employees with a contract (*insurance, social security, union membership*)	Number of public (private) paid employees have contract (*insurance, social security, union membership)/ Total number of public (private) paid employees*

Definitions for these variables are broadly applicable to apply to workers across industries and sectors. However, the lack of specificity should be noted such that “contracts” or “insurance” may not be externally consistent except in the broadest terms. Since these survey questions vary from country to country, there could be variation in question wording and the local understanding of terms, in addition to the assumption that the question is included at all. There is no indication that these terms are based on an internationally accepted concept (such as with the occupational classifications). Thus, there is a smaller sample of countries with this information. The definition of each benefits is outlined as follows:

Table 23: I2D2 Benefits and definitions

Variable	*Description*
Work contract	*Indicates if a person has a signed (formal) contract, regardless of duration.*
Social Security	*constructed in the presence of an explicit question on pension plans or social security.*
Health insurance	*constructed if there is an explicit question about health security.*
Union membership	*constructed if there is an explicit question about trade unions.*

7.15 Relative wages of public sector employees

Two additional sets of indicators on pay compression ratios in the public sector were included in the Version 2.0 of the dataset. Both indicators constructed using data from the 2017 cycle of the International Comparison Program's (ICP) "Data for Researchers" dataset. The ICP is a worldwide statistical initiative to collect and compile comparable price and national accounts expenditure data and estimate purchasing power parties (PPPs) for the world's economies. The program is implemented as a global partnership of national and regional agencies and managed by the ICP Global Office at the World Bank, under the auspices of the United Nations Statistical Commission (UNSC). More information is available on the Program's website at https://www.worldbank.org/en/programs/icp/.

Table 24: WWBI Relative wage variables

Variable Name	***Description***
Relative wage of Senior Officials (Professionals, Technicians) in public (private) sector	*Average weekly wage of Senior Officials (Professionals, Technicians) in public (private) sector/ Average weekly wage of clerk in public (private) sector.* *Occupation categories adhere to the ILO's International Standard of Classification of Occupations using the 1988 nomenclature (ISCO-88).*
Wage compression ratio in public sector, by occupation (using clerical occupations as reference)	*The ratio of the wages of the indexed occupations to all clerical occupations (as the benchmark category).* ***Indicator group 1:*** • *Pay compression ratio in public sector, by occupation: Senior officials (clerk as reference)* • *Pay compression ratio in public sector, by occupation: Judge (clerk as reference)* • *Pay compression ratio in public sector, by occupation: Government economist (clerk as reference)* • *Pay compression ratio in public sector, by occupation: Hospital doctor (clerk as reference)* • *Pay compression ratio in public sector, by occupation: Hospital nurse (clerk as reference)* • *Pay compression ratio in public sector, by occupation: University teacher (clerk as reference)* • *Pay compression ratio in public sector, by occupation: Secondary school teacher (clerk as reference)* • *Pay compression ratio in public sector, by occupation: Primary school teacher (clerk as reference)* • *Pay compression ratio in public sector, by occupation: Police officer (clerk as reference)*
Cross-country public sector wage comparison ratio, by occupation (using mean)	*The ratio of the wages of the indexed occupations within the reference country to the global mean for the same category. In order to control for differences in prices across nations, the administrative data in local currencies are converted and benchmarked in US$ using PPP*

	Conversion Factors (Household Expenditure) to allow adequate comparisons. ***Indicator group 2:*** • *Cross-country public sector pay comparison ratio, by occupation: Senior official (using mean)* • *Cross-country public sector pay comparison ratio, by occupation: Judge (using median)* • *Cross-country public sector pay comparison ratio, by occupation: Government economist (using mean)* • *Cross-country public sector pay comparison ratio, by occupation: Hospital doctor (using mean)* • *Cross-country public sector pay comparison ratio, by occupation: Hospital nurse (using mean)* • *Cross-country public sector pay comparison ratio, by occupation: University teacher (using mean)* • *Cross-country public sector pay comparison ratio, by occupation: Secondary school teacher (using mean)* • *Cross-country public sector pay comparison ratio, by occupation: Primary school teacher (using mean)* • *Cross-country public sector pay comparison ratio, by occupation: Police officer (using mean)*
Cross-country public sector wage comparison ratio, by occupation (using median)	*The ratio of the wages of the indexed occupations within the reference country to the global median for the same category. In order to control for differences in prices across nations, the administrative data in local currencies are converted and benchmarked in US$ using PPP Conversion Factors (Household Expenditure) to allow adequate comparisons.* ***Indicator group 2:*** • *Cross-country public sector pay comparison ratio, by occupation: Senior official (using median)* • *Cross-country public sector pay comparison ratio, by occupation: Judge (using median)* • *Cross-country public sector pay comparison ratio, by occupation: Government economist (using median)* • *Cross-country public sector pay comparison ratio, by occupation: Hospital doctor (using median)* • *Cross-country public sector pay comparison ratio, by occupation: Hospital nurse (using median)* • *Cross-country public sector pay comparison ratio, by occupation: University teacher (using median)* • *Cross-country public sector pay comparison ratio, by occupation: Secondary school teacher (using median)* • *Cross-country public sector pay comparison ratio, by occupation: Primary school teacher (using median)* • *Cross-country public sector pay comparison ratio, by occupation: Police officer (using median)*

7.16 Wage Bill

Indicators on the relative size of the public sector wage bill are sources from the IMF Government Compensation and Employment Dataset, 2022. The wage bill is defined as the total compensation (in cash or in-kind) payable to a government employee in exchange for work. Wage bill includes wages and salaries, allowances, and social security contributions made on behalf of employees to social insurance schemes (IMF 2014). The IMF provides a detailed explanation of the construction of the indicators and the suggested caveats for using the cross-country analysis data (IMF 2022). It notes that the measurement of the wage bill might differ depending on the coverage, definitions, and different ways of public service provision:

- The base of the wage bill may vary considerably across countries. While high-income countries prefer expressing wage expenditure on an accrual basis (including an imputation for the difference between the current period pensions and the contributions paid for these benefits), the rest prefers cash.
- Recording of benefits and bonuses (sometimes recorded as expenditure in goods and services), unit of measurements (number of employees vs. full-time equivalents), and the definition of employment (permanent vs. temporary employees) might essentially lead to different wage bill calculations.
- Issues of comparability could also arise by the different ways in which governments provide public services. For example, in France, most health care professionals are government employees. While in the Netherlands, they are contractors whose compensation is classified under Goods and Service expenditure instead of Compensation of Employee.

Table 25: WWBI Wage bill variables

Variable Name	***Source***	***Description***
General government wage bill as a percentage of GDP	*IMF*	*The data sources used to construct the harmonized series of wage bill are IMF's World Economic Outlook (WEO) and Government Finance Statistics (GFS), OECD's General Government Accounts, and EUROSTAT's Annual Government Finance Statistics and AMECO. For earlier data for a few countries the main source is Cusack, Thomas R., 2006, "Public Finance Data for 20 OECD Countries." For the harmonized government employment series, the main sources include the ILO's LABORSTA data (public sector employment and employment of general government sector), ILOSTAT (employment by institutional sector), as well as data from individual countries.*
General government wage bill as a percentage of general government expenditures		

8. Appendices

A1: Data filters

The main aim of WWBI is to provide information on the public sector labor market using household surveys. To avoid possible bias due to missing values, we created data quality filters to help screen surveys with more than 40 percent of missing values in our core variables: employment, wages, age, gender, and education-related variables. In the development of the filters, a hierarchical process was used that follows the order of the filters as described below (employment, wages, socio-demographic, and education). Once a survey fails to fulfill any of the conditions imposed, it gets a marked with the filter, and on all the subsequent ones. In the final dataset, surveys that did not pass the employment filters are removed since these are essential for identifying the sector of employment were excluded when calculating indicators for the WWBI. In addition, we also created quality filters to ensure the surveys used to construct WWBI are representative of the labor market. Based on which variables the filters are addressing, we divide the filters into four categories. The detailed filters are listed below:

Data Quality Filters

1. Filters for employment variables (used to exclude entire surveys from the dataset)

a. *Survey does not include labor variables (f01 to f03)*

f1. Survey does not have labor status for anyone:

missing_lstatus ==1

f2. Survey does not have employment status for employed individuals:

m_empstat ==1 if m_lstatus ==0

f3. Survey does not list sector of activity for people currently working:

m_ocusec ==1 if m_lstatus == 0

b. *More than 40 percent of data is missing (f11 to f13)*

f11. More than 40% of the data (observations) on labor status are missing:

m_lstatus > 0.4 if m_lstatus ~= 1

f12. More than 40% of the data on employment status are missing:

m_empstat > 0.4 if m_lstatus ~=.

f13. More than 40% of the data on occupation are missing:

mpubsec_empl1 > 0.4 if mpubsec_empl1 ~=.

c. *Other employment filters (f5b to f5d)*

f5b. All observations are employed:

lstatus1 == 1

f5c. All observations are employees (i.e., no self-employers or employers)

Empstat3 ==. & empstat4 ==.

f5d. All paid employees are either all public or all private:

pubsec_empl1 == 1 | pubsec_empl == 0 if pubsec_empl ~=.

2. Filters for missing wages (used to exclude wage related indicators from dataset)

1. More than 40% missing observations for wages of paid employees:

m_wpm>0.4 | m_wpm==.

f6a. More than 40% missing observations for wage unit:

m_unitwage > 0.4 | m_unitwage==.

3. Filters for missing age, gender, or location (used to exclude demographic indicators from dataset)

f7. More than 40% missing observations for age, gender, or location:

m_age > 0.4 | m_gender > 0.4 | m_urb > 0.4

4. Filters for missing education (used to exclude education related indicators from dataset)

f8. More than 40% missing observations for education qualifications achieved:

m_edulevel3 > 0.4

Additional filters for exclusion of surveys

1. Filters for small size of public sector (used to exclude survey from dataset)

a. Public sector paid employment (share of total) is less than 5% of employed:

ps1 < 0.05

b. Public sector paid employment (share of total) is less than 200 individuals:

obs_pub < 200

2. Filters for only urban surveys (used to exclude survey from dataset)

a. Survey includes only urban sample:

mean_urban == 1

3. Filters for multiple surveys for the same country and year (used to exclude survey from dataset)

a. Identify and drop duplicate surveys:

bysort country year: gen dup = cond(_N==1,0,n)

drop if dup > 0

A detailed description of the number of surveys that passed many surveys passed all filters and how many surveys failed each filter are listed below. 669 surveys were excluded from our analysis as they didn't pass the employment filter. For 243 surveys that didn't pass wage, sociodemographic and education filters, only those statistics were estimated that do not utilize the variables included within these filters. For 886 surveys that passed all the filters, all estimates were constructed.

A2: Outliers in the I2D2 dataset

Data used in the estimation of the Worldwide Bureaucracy indicators are collected by national or regional statistical agencies using a probability-based sampling frame. Each individual in the data represents a number of individuals in the population, and the rate of representation of the individual in the sample to the population is given by the weight. In the construction of the weights, however, this representativeness is designed, based on a handful of key traits such as age, sex, household size and composition, education received, and perhaps race and ethnicity to ensure that sample averages correspond to population averages. Similarly, in the case for Labor Force Surveys, close attention is paid to ensuring that labor force participation and unemployment rates closely align with the population. Questions that do not have specific weights assigned to them, rely on the statistical agency's ability to survey a large-enough number of people to reduce the

likelihood of statistical skewedness. Still, in the process of construction of the Worldwide Bureaucracy Indicators, we find certain indicators to be statistical outliers.

1 Outlier observations were identified for all variables using a three-step process; by looking at variations over year within (1) country, (2) their World Bank defined regions, and (3) their income category. For the first step, we calculated the mean and standard deviation of the indicator across years for each country that had at least four years of observations. This allows us to identify one-off statistical anomalies where the indicator sharply jumps over three standard deviations from the mean and is not explained by a localized trend or accompanying literature.

a. by country: variable – (mean_variable ± (3 * sd_variable)) if ccode_obs > 3

This method is not suitable for countries with 3 or fewer observations since the standard deviation mechanically stretches to accommodate the outlies. In order to circumvent this problem, observations that represented significant structural breaks from an overall country trend were identifies using the following method. Here, outliers are identified as observations where the difference between them and the country's average is greater than 20 percent of the entire variance of the indicator. This allows for the identification of outliers, even for countries with fewer than 4 observations by circumventing the need to use the standard deviation.

b. by country: variable – mean_variable > (max(variable) – min(variable))

2 Additionally, for countries with fewer than 4 observation, outliers were also identified using their deviations from the mean of their region and income categories – as defined by the World Bank. For these countries, we rely on the variation of the entire sample of countries for the entire set of years. Outliers are identified if they jump over a quarter of this entire variation. The same method is applied for the regional and income classifications which allow us to identify additional outliers and structural breaks that were introducing noise within the dataset.

c. by income_group: variable – (mean_variable ± (3 * sd_variable)) if ccode_obs < 4

d. by region: variable – (mean_variable ± (3 * sd_variable)) if ccode_obs < 4

This method was applied to all indicators except the wage bill indicators and the p-values associated with the wage regressions. This code along with the cleaning and estimation .do files and supplementary materials have been archived on GitHub here.

A3: Explanatory note on wage premium transformation

Since the untransformed $\hat{\beta}$ give is merely an approximation of the actual premium and the discrepancy becomes larger when the $\hat{\beta}$ > ± 20%. In our sample, we calculate 7497 premiums for our 13 indicators across all countries and years. A quick comparison of the untransformed ($\hat{\beta}$) and transformed ($e^{\hat{\beta}} - 1$) reveal that while over 85 percent of reported premia are lower than this threshold. Even this transformation meaningfully affects a relatively small portion of indicators, the untransformed premiums are less precise, and we instead report the transformed versions as follows

This is the equation we estimate is

$$\log y = \beta_0 + \beta_1 * x \quad \textbf{(1)}$$

Which can also be written as:

$$e^{\log y} = e^{(\beta_0 + \beta_1 * x)} \quad \textbf{(2)}$$

The coefficient of interest is the change in the percentage change in y brought by a change in x by 1 unit. We can express the change in y as:

$$\Delta y = e^{(\beta_0 + \beta_1 (x_0 + \Delta x))} - e^{(\beta_0 + \beta_1 (x_0))} \quad \textbf{(3)}$$

Simplifying equation 3 leaves us with:

$$\Delta y = y_0 (e^{\beta_1 \Delta x} - 1) \quad \textbf{(4)}$$

A4: Full list of surveys in WWBI with original data source

Table 26: Complete WWBI 3.1 Survey list

Country Name	***Survey Name***
Afghanistan	2003 National Risk and Vulnerability Assessment, Living Conditions Survey
Afghanistan	2007 National Risk and Vulnerability Assessment, Living Conditions Survey
Afghanistan	2011 National Risk and Vulnerability Assessment, Living Conditions Survey
Afghanistan	2013 Afghanistan Living Conditions Survey
Angola	2000 Household Income and Expenditure Survey (IDR)
Angola	2008 Multiple Industry Cluster Survey (MICS)
Angola	2014 Angola Population Census
Albania	2002 Living Standards Measurement Survey
Albania	2004 Living Standards Measurement Survey
Albania	2004 Living Standards Measurement Survey
Albania	2008 Living Standards Measurement Survey
Albania	2012 Living Standards Measurement Survey
Argentina	2000 Encuesta Permanente de Hogares Contínua
Argentina	2001 Encuesta Permanente de Hogares Contínua
Argentina	2002 Encuesta Permanente de Hogares Contínua
Argentina	2003 Encuesta Permanente de Hogares Contínua
Argentina	2004 Encuesta Permanente de Hogares Contínua
Argentina	2005 Encuesta Permanente de Hogares Contínua
Argentina	2006 Encuesta Permanente de Hogares Contínua
Argentina	2007 Encuesta Permanente de Hogares Contínua
Argentina	2008 Encuesta Permanente de Hogares Contínua
Argentina	2009 Encuesta Permanente de Hogares Contínua
Argentina	2010 Encuesta Permanente de Hogares Contínua
Argentina	2011 Encuesta Permanente de Hogares Contínua
Argentina	2012 Encuesta Permanente de Hogares Contínua
Argentina	2013 Encuesta Permanente de Hogares Contínua
Argentina	2014 Encuesta Permanente de Hogares Contínua
Argentina	2017 Encuesta Permanente de Hogares Contínua
Argentina	2018 Encuesta Permanente de Hogares Contínua
Argentina	2019 Encuesta Permanente de Hogares Contínua
Argentina	2020 Encuesta Permanente de Hogares Contínua
Armenia	2008 Armenia Integrated Living Conditions Survey
Armenia	2011 Armenia Integrated Living Conditions Survey
Australia	2009 Household, Income and Labor Dynamics in Australia

Australia	2010 Household, Income and Labor Dynamics in Australia
Australia	2011 Household, Income and Labor Dynamics in Australia
Australia	2012 Household, Income and Labor Dynamics in Australia
Australia	2013 Household, Income and Labor Dynamics in Australia
Australia	2014 Household, Income and Labor Dynamics in Australia
Australia	2015 Household, Income and Labor Dynamics in Australia
Azerbaijan	2011 Azerbaijan Household Survey on Monitoring Targeted Social Assistance Programme
Azerbaijan	2015 Azerbaijan Monitoring Survey of Social Welfare
Benin	2018 Integrated Modular Survey on Households Living Conditions
Burkina Faso	2003 Enquête burkinabé sur les conditions de vie des ménages
Burkina Faso	2009 Enquete Integrale sur les Conditions de Vie des Menages
Burkina Faso	2014 Enquête Multisectorielle Continue
Bangladesh	2005 Labor Force Survey
Bangladesh	2010 Labor Force Survey
Bangladesh	2013 Labor Force Survey
Bangladesh	2015 Labor Force Survey
Bosnia and Herzegovina	2001 Living Standards Measurement Survey
Bosnia and Herzegovina	2006 Labor Force Survey
Bosnia and Herzegovina	2007 Labor Force Survey
Bosnia and Herzegovina	2008 Labor Force Survey
Bosnia and Herzegovina	2009 Labor Force Survey
Bosnia and Herzegovina	2010 Labor Force Survey
Bosnia and Herzegovina	2011 Labor Force Survey
Bosnia and Herzegovina	2012 Labor Force Survey
Bosnia and Herzegovina	2013 Labor Force Survey
Bosnia and Herzegovina	2014 Labor Force Survey
Bosnia and Herzegovina	2015 Labor Force Survey
Bosnia and Herzegovina	2016 Labor Force Survey
Bosnia and Herzegovina	2017 Labor Force Survey
Bosnia and Herzegovina	2018 Labor Force Survey

Bosnia and Herzegovina	2019 Labor Force Survey
Bolivia	2000 Encuesta Continua de Hogares
Bolivia	2001 Encuesta de Hogares
Bolivia	2002 Encuesta de Hogares
Bolivia	2003 Encuesta de Hogares
Bolivia	2005 Encuesta de Hogares
Bolivia	2006 Encuesta de Hogares
Bolivia	2007 Encuesta de Hogares
Bolivia	2008 Encuesta de Hogares
Bolivia	2009 Encuesta de Hogares
Bolivia	2011 Encuesta de Hogares
Bolivia	2012 Encuesta de Hogares
Bolivia	2014 Encuesta de Hogares
Bolivia	2015 Encuesta de Hogares
Bolivia	2016 Encuesta de Hogares
Bolivia	2017 Encuesta de Hogares
Bolivia	2020 Encuesta de Hogares
Brazil	2001 Pesquisa Nacional por Amostra de Domicilios
Brazil	2002 Pesquisa Nacional por Amostra de Domicilios
Brazil	2003 Pesquisa Nacional por Amostra de Domicilios
Brazil	2004 Pesquisa Nacional por Amostra de Domicilios
Brazil	2005 Pesquisa Nacional por Amostra de Domicilios
Brazil	2006 Pesquisa Nacional por Amostra de Domicilios
Brazil	2007 Pesquisa Nacional por Amostra de Domicilios
Brazil	2008 Pesquisa Nacional por Amostra de Domicilios
Brazil	2009 Pesquisa Nacional por Amostra de Domicilios
Brazil	2011 Pesquisa Nacional por Amostra de Domicilios
Brazil	2012 Pesquisa Nacional por Amostra de Domicilios
Brazil	2014 Pesquisa Nacional por Amostra de Domicilios
Brazil	2015 Pesquisa Nacional por Amostra de Domicilios
Brazil	2016 Pesquisa Nacional por Amostra de Domicilios
Bhutan	2003 Bhutan Living Standard Survey
Bhutan	2007 Bhutan Living Standard Survey
Bostwana	2009 The Botswana Core Welfare Indicators Survey
Bostwana	2015 Botswana Multi-Topic Household Survey
Central African Republic	2008 Enquête Centrafricaine pour le Suivi – Evaluation du Bien – Etre
Chile	2000 Chile National Socioeconomic Characterization Survey
Chile	2003 Chile National Socioeconomic Characterization Survey
Chile	2006 Chile National Socioeconomic Characterization Survey
Chile	2009 Chile National Socioeconomic Characterization Survey
Chile	2011 Chile National Socioeconomic Characterization Survey

Chile	2013 Chile National Socioeconomic Characterization Survey
Chile	2015 Chile National Socioeconomic Characterization Survey
Chile	2017 Chile National Socioeconomic Characterization Survey
Chile	2020 Chile National Socioeconomic Characterization Survey
China	2002 Chinese Household Income Project
China	2013 Chinese Household Income Project
Cameroon	2001 Enquête Camerounaise Auprès des Ménages
Cameroon	2007 Enquête Camerounaise Auprès des Ménages
Cameroon	2010 Employment and Informal Sector Survey
Cameroon	2014 Enquête Camerounaise Auprès des Ménages
Cameroon	2021 Enquête Camerounaise Auprès des Ménages
Dem. Rep. of Congo	2004 Survey on Employment, the Informal Sector and Household Living Conditions
Dem. Rep. of Congo	2012 Survey on Employment, the Informal Sector and Household Living Conditions
Congo	2005 Congo Le Questionnaire des Indicateurs de Base de Bien-être
Congo	2011 Congo Enquête Congolaise auprès des Ménages
Colombia	2001 Encuesta Continua de Hogares
Colombia	2002 Encuesta Continua de Hogares
Colombia	2003 Encuesta Continua de Hogares
Colombia	2004 Encuesta Continua de Hogares
Colombia	2005 Encuesta Continua de Hogares
Colombia	2006 Gran Encuesta Integrada de Hogares
Colombia	2007 Gran Encuesta Integrada de Hogares
Colombia	2008 Gran Encuesta Integrada de Hogares
Colombia	2009 Gran Encuesta Integrada de Hogares
Colombia	2010 Gran Encuesta Integrada de Hogares
Colombia	2011 Gran Encuesta Integrada de Hogares
Colombia	2012 Gran Encuesta Integrada de Hogares
Colombia	2013 Gran Encuesta Integrada de Hogares
Colombia	2014 Gran Encuesta Integrada de Hogares
Colombia	2015 Gran Encuesta Integrada de Hogares
Colombia	2016 Gran Encuesta Integrada de Hogares
Colombia	2017 Gran Encuesta Integrada de Hogares
Comoros	2004 Enquête Intégrale auprès des Ménages
Comoros	2013 Enquête Intégrale auprès des Ménages
Cabo Verde	2007 Core Welfare Indicator Questionnaire Survey
Costa Rica	2000 Encuesta de Hogares de Propósitos Múltiples
Costa Rica	2001 Encuesta de Hogares de Propósitos Múltiples
Costa Rica	2002 Encuesta de Hogares de Propósitos Múltiples
Costa Rica	2003 Encuesta de Hogares de Propósitos Múltiples
Costa Rica	2004 Encuesta de Hogares de Propósitos Múltiples

Costa Rica	2005 Encuesta de Hogares de Propósitos Múltiples
Costa Rica	2006 Encuesta de Hogares de Propósitos Múltiples
Costa Rica	2007 Encuesta de Hogares de Propósitos Múltiples
Costa Rica	2008 Encuesta de Hogares de Propósitos Múltiples
Costa Rica	2009 Encuesta de Hogares de Propósitos Múltiples
Costa Rica	2010 Encuesta Nacional de Hogares
Costa Rica	2012 Encuesta Nacional de Hogares
Costa Rica	2015 Encuesta Nacional de Hogares
Costa Rica	2016 Encuesta Nacional de Hogares
Costa Rica	2018 Encuesta Continua de Empleo
Costa Rica	2019 Encuesta Continua de Empleo
Costa Rica	2020 Encuesta Continua de Empleo
Djibouti	2002 Household Survey for Social Indicators
Djibouti	2015 Enquête sur l'emploi, le secteur informel et la consommation des ménages
Dominican Republic	2000 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2001 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2002 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2003 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2004 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2005 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2006 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2007 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2008 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2009 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2010 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2011 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2012 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2013 Encuesta Nacional de la fuerza de trabajo
Dominican Republic	2015 Encuesta Nacional de la fuerza de trabajo
Ecuador	2000 Encuesta Periodica sobre Empleo y Desempleo
Ecuador	2003 Encuesta Nacional de Empleo, Desempleo y Subempleo

Ecuador	2004 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2005 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2006 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2007 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2008 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2009 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2010 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2011 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2012 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2013 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2014 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2015 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2016 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2017 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2019 Encuesta Nacional de Empleo, Desempleo y Subempleo
Ecuador	2020 Encuesta Nacional de Empleo, Desempleo y Subempleo
Egypt	2004 Egypt Household Income, Expenditure and Consumption Survey
Egypt	2005 Egypt Labor Market Panel Survey
Egypt	2015 Egypt Household Income, Expenditure and Consumption Survey
Ethiopia	2000 Welfare Monitoring Survey
Ethiopia	2003 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2004 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2005 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2006 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2009 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2010 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2011 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2012 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2013 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2014 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2015 Ethiopian Urban Employment Unemployment Survey
Ethiopia	2016 Ethiopian Urban Employment Unemployment Survey
Micronesia	2005 Household Income and Expenditure Survey
Gabon	2005 Enquête Gabonaise pour l'Evaluation et le Suivi de la Pauvreté
Gabon	2017 Gabon: Enquête Gabonaise pour l'Evaluation et le Suivi de la Pauvreté
Georgia	2008 Household Income Survey
Georgia	2009 Household Income Survey
Georgia	2011 Household Income Survey
Georgia	2012 Household Income Survey
Georgia	2013 Household Income Survey
Ghana	2005 Living Standards Survey

Ghana	2012 Living Standards Survey
Ghinea	2002 Core Welfare Indicators Survey
Ghinea	2007 Core Welfare Indicators Survey
Ghinea	2012 Enquête Légère pour l'Evaluation de la Pauvreté
Gambia	2010 Gambia Integrated Household Survey
Gambia	2015 Gambia Intgrated Household Survey
Guinea-Bissau	2010 Poverty Assessment Survey
Guinea-Bissau	2018 Household Income and Expenditure Survey
Guatemala	2000 Encuesta Nacional de Condiciones de Vivienda
Guatemala	2002 Encuesta Nacional de empleo e ingresos
Guatemala	2003 Encuesta Nacional de empleo e ingresos
Guatemala	2004 Encuesta Nacional de empleo e ingresos
Guatemala	2006 Encuesta Nacional de Condiciones de Vivienda
Guatemala	2011 Encuesta Nacional de Condiciones de Vivienda
Guatemala	2017 Encuesta Nacional de empleo e ingresos
Guatemala	2018 Encuesta Nacional de empleo e ingresos
Guatemala	2019 Encuesta Nacional de empleo e ingresos
Honduras	2001 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2002 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2003 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2004 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2005 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2006 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2007 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2008 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2009 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2010 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2011 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2012 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2013 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2014 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2015 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2016 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2017 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2018 Encuesta Permanente de Hogares y Propositos Multiples
Honduras	2019 Encuesta Permanente de Hogares y Propositos Multiples
Haiti	2001 Enquête sur les Conditions de Vie en Haïti
Hungary	2004 Household Budget Survey
Hungary	2005 Household Budget Survey
India	2005 India Human Development Survey
Iran	2019 Household Expenditure and Income Survey
Jamaica	2001 Jamaica Survey of Living Conditions

Jamaica	2002 Jamaica Survey of Living Conditions
Jordan	2002 Household Income and Expenditure Survey
Jordan	2007 Labor Force Survey
Jordan	2008 Labor Force Survey
Jordan	2009 Labor Force Survey
Jordan	2010 Labor Force Survey
Jordan	2011 Labor Force Survey
Jordan	2012 Labor Force Survey
Jordan	2013 Labor Force Survey
Jordan	2014 Labor Force Survey
Jordan	2016 Labor Force Survey
Kazakhstan	2006 Household Budget Survey
Kazakhstan	2010 Household Budget Survey
Kazakhstan	2013 Household Budget Survey
Kenya	2005 Household Budget Survey
Kenya	2015 Kenya Integrated Household Budget Survey
Cambodia	2003 Cambodia Socio-Economic Survey
Cambodia	2006 Cambodia Socio-Economic Survey
Cambodia	2008 Cambodia Socio-Economic Survey
Cambodia	2009 Cambodia Socio-Economic Survey
Cambodia	2012 Cambodia Labor Force Survey
Lebanon	2011 Labor Force Survey
Lebanon	2018 Labor Force Survey
Liberia	2010 Labor Force Survey
Sri Lanka	2000 Labor Force Survey
Sri Lanka	2001 Labor Force Survey
Sri Lanka	2002 Labor Force Survey
Sri Lanka	2003 Labor Force Survey
Sri Lanka	2004 Labor Force Survey
Sri Lanka	2006 Labor Force Survey
Sri Lanka	2007 Labor Force Survey
Sri Lanka	2008 Labor Force Survey
Sri Lanka	2009 Household Income and Expenditure Survey
Sri Lanka	2011 Labor Force Survey
Sri Lanka	2012 Labor Force Survey
Sri Lanka	2013 Labor Force Survey
Sri Lanka	2014 Labor Force Survey
Sri Lanka	2015 Labor Force Survey
Sri Lanka	2016 Household Income and Expenditure Survey
Lesotho	2017 Household Budget Survey
Morocco	2005 Morocco ENSLE Survey
Morocco	2006 Morocco ENSLE Survey

Morocco	2007 Morocco ENSLE Survey
Morocco	2008 Morocco ENSLE Survey
Morocco	2009 Morocco ENSLE Survey
Moldova	2006 Labor Force Survey
Moldova	2007 Labor Force Survey
Moldova	2008 Labor Force Survey
Moldova	2009 Labor Force Survey
Moldova	2010 Labor Force Survey
Moldova	2011 Labor Force Survey
Moldova	2012 Labor Force Survey
Moldova	2014 Labor Force Survey
Moldova	2015 Labor Force Survey
Madagascar	2001 Enquête Intégrale sur les Conditions de Vie des Ménages
Madagascar	2005 Madagascar Enquêtes Permanentes/Périodiques auprès des. Ménages
Madagascar	2012 Enquête nationale sur le suivi des objectifs du millénaire pour le développement
Maldives	2002 Household Income and Expenditure Survey
Maldives	2004 Household Income and Expenditure Survey
Maldives	2009 Household Income and Expenditure Survey
Maldives	2016 Household Income and Expenditure Survey
Maldives	2019 Household Income and Expenditure Survey
Mexico	2000 Encuesta Nacional de Ingresos y Gastos de los Hogares
Mexico	2002 Encuesta Nacional de Ingresos y Gastos de los Hogares
Mexico	2005 Encuesta Nacional de Ocupacion y Empleo
Mexico	2006 Encuesta Nacional de Ocupacion y Empleo
Mexico	2007 Encuesta Nacional de Ocupacion y Empleo
Mexico	2008 Encuesta Nacional de Ingresos y Gastos de los Hogares
Mexico	2009 Encuesta Nacional de Ocupacion y Empleo
Mexico	2010 Encuesta Nacional de Ingresos y Gastos de los Hogares
Mexico	2011 Encuesta Nacional de Ocupacion y Empleo
Mexico	2012 Encuesta Nacional de Ingresos y Gastos de los Hogares
Mexico	2013 Encuesta Nacional de Ocupacion y Empleo
Mexico	2014 Encuesta Nacional de Ocupacion y Empleo
Mexico	2015 Encuesta Nacional de Ocupacion y Empleo
Mexico	2016 Encuesta Nacional de Ocupacion y Empleo
Mexico	2018 Encuesta Nacional de Ocupacion y Empleo
Mexico	2019 Encuesta Nacional de Ocupacion y Empleo
Mali	2003 Integrated Household Survey
Mali	2010 Enquête Permanente Auprès des Ménages
Mali	2018 Household Budget Survey
Montenegro	2010 Labor Force Survey
Montenegro	2011 Labor Force Survey

Mongolia	2002 Household Income and Expenditure Survey
Mongolia	2006 Labor Force Survey
Mongolia	2010 Household Socio-Economic Survey
Mongolia	2011 Household Socio-Economic Survey
Mongolia	2014 Labor Force Survey
Mongolia	2019 Labor Force Survey
Mongolia	2020 Labor Force Survey
Mongolia	2021 Labor Force Survey
Mozambique	2002 Household Budget Survey
Mozambique	2008 Household Budget Survey
Mozambique	2012 National Survey on the Causes of Death
Mozambique	2014 Mozambique Orcamento Familiar (IOF) Survey
Mauritania	2000 Enquête Permanente sur les Conditions de Vie des ménages
Mauritania	2004 Enquête Permanente sur les Conditions de Vie des ménages
Mauritania	2008 Enquête Permanente sur les Conditions de Vie des ménages
Mauritania	2014 Enquête Permanente sur les Conditions de Vie des ménages
Mauritius	2001 Continuos Multi-Purpose Household Survey
Mauritius	2004 Continuos Multi-Purpose Household Survey
Mauritius	2005 Continuos Multi-Purpose Household Survey
Mauritius	2006 Continuos Multi-Purpose Household Survey
Mauritius	2007 Continuos Multi-Purpose Household Survey
Mauritius	2008 Continuos Multi-Purpose Household Survey
Mauritius	2009 Continuos Multi-Purpose Household Survey
Mauritius	2010 Continuos Multi-Purpose Household Survey
Mauritius	2011 Continuos Multi-Purpose Household Survey
Mauritius	2012 Continuos Multi-Purpose Household Survey
Malawi	2004 Integrated Household Survey
Malawi	2019 Antegrated Household Survey
Namibia	2012 Labor Force Survey
Namibia	2013 Labor Force Survey
Namibia	2014 Labor Force Survey
Namibia	2015 Household Income and Expenditure Survey
Nigeria	2003 Living Standards Survey
Nigeria	2009 Living Standards Survey
Nicaragua	2001 Encuesta de Medicion del Nivel de Vida
Nicaragua	2005 Encuesta de Medicion del Nivel de Vida
Nepal	2008 Labor Force Survey
Nepal	2013 Annual Household Survey
Nepal	2014 Annual Household Survey
Pakistan	2001 Labor Force Survey
Pakistan	2003 Labor Force Survey
Pakistan	2004 Pakistan Social and Living Standards Measurement

Pakistan	2005 Labor Force Survey
Pakistan	2006 Pakistan Social and Living Standards Measurement
Pakistan	2008 Labor Force Survey
Pakistan	2009 Pakistan Social and Living Standards Measurement
Pakistan	2010 Labor Force Survey
Pakistan	2012 Labor Force Survey
Pakistan	2013 Labor Force Survey
Pakistan	2014 Labor Force Survey
Panama	2000 Encuesta de Hogares
Panama	2001 Encuesta de Hogares
Panama	2002 Encuesta de Hogares
Panama	2003 Encuesta de Hogares
Panama	2004 Encuesta de Hogares
Panama	2005 Encuesta de Hogares
Panama	2006 Encuesta de Hogares
Panama	2007 Encuesta de Hogares
Panama	2008 Encuesta de Hogares
Panama	2009 Encuesta de Hogares
Panama	2010 Encuesta de Hogares
Panama	2011 Encuesta de Hogares
Panama	2012 Encuesta de Hogares
Panama	2015 Encuesta de Hogares
Panama	2016 Encuesta de Hogares
Panama	2017 Encuesta de Hogares
Panama	2018 Encuesta de Hogares
Peru	2000 Encuesta Nacional de Hogares
Peru	2001 Encuesta Nacional de Hogares
Peru	2002 Encuesta Nacional de Hogares
Peru	2004 Encuesta Nacional de Hogares
Peru	2005 Encuesta Nacional de Hogares
Peru	2006 Encuesta Nacional de Hogares
Peru	2007 Encuesta Nacional de Hogares
Peru	2008 Encuesta Nacional de Hogares
Peru	2009 Encuesta Nacional de Hogares
Peru	2010 Encuesta Nacional de Hogares
Peru	2011 Encuesta Nacional de Hogares
Peru	2012 Encuesta Nacional de Hogares
Peru	2013 Encuesta Nacional de Hogares
Peru	2014 Encuesta Nacional de Hogares
Peru	2015 Encuesta Nacional de Hogares
Peru	2016 Encuesta Nacional de Hogares
Peru	2018 Encuesta Nacional de Hogares

Peru	2019 Encuesta Nacional de Hogares
Philippines	2000 Labor Force Survey
Philippines	2001 Labor Force Survey
Philippines	2002 Labor Force Survey
Philippines	2003 Labor Force Survey
Philippines	2004 Labor Force Survey
Philippines	2005 Labor Force Survey
Philippines	2006 Labor Force Survey
Philippines	2007 Labor Force Survey
Philippines	2008 Labor Force Survey
Philippines	2009 Labor Force Survey
Philippines	2010 Labor Force Survey
Philippines	2011 Labor Force Survey
Philippines	2012 Labor Force Survey
Philippines	2013 Labor Force Survey
Philippines	2014 Labor Force Survey
Philippines	2015 Labor Force Survey
Philippines	2016 Labor Force Survey
Philippines	2017 Labor Force Survey
Philippines	2018 Labor Force Survey
Philippines	2019 Labor Force Survey
Palau	2000 Population Census
Papua New Guinea	2009 Household Income and Expenditure Survey
Poland	2001 Household Budget Survey
Poland	2002 Household Budget Survey
Poland	2003 Household Budget Survey
Poland	2004 Household Budget Survey
Puerto Rico	2000 Census
Puerto Rico	2005 Census
Paraguay	2001 Encuesta Integrada de Hogares
Paraguay	2002 Encuesta Permanente de Hogares
Paraguay	2003 Encuesta Permanente de Hogares
Paraguay	2004 Encuesta Permanente de Hogares
Paraguay	2005 Encuesta Permanente de Hogares
Paraguay	2006 Encuesta Permanente de Hogares
Paraguay	2007 Encuesta Permanente de Hogares
Paraguay	2008 Encuesta Permanente de Hogares
Paraguay	2009 Encuesta Permanente de Hogares
Paraguay	2010 Encuesta Permanente de Hogares
Paraguay	2011 Encuesta Permanente de Hogares
Paraguay	2012 Encuesta Permanente de Hogares

Paraguay	2013 Encuesta Permanente de Hogares
Paraguay	2014 Encuesta Permanente de Hogares
Paraguay	2015 Encuesta Permanente de Hogares
Paraguay	2016 Encuesta Permanente de Hogares
Paraguay	2017 Encuesta Permanente de Hogares
Paraguay	2019 Encuesta Permanente de Hogares
Paraguay	2020 Encuesta Permanente de Hogares
Russia	2000 Russian Longitudinal Monitoring Survey
Russia	2001 Russian Longitudinal Monitoring Survey
Russia	2002 Russian Longitudinal Monitoring Survey
Russia	2003 Russian Longitudinal Monitoring Survey
Russia	2004 Russian Longitudinal Monitoring Survey
Russia	2005 Russian Longitudinal Monitoring Survey
Russia	2006 Russian Longitudinal Monitoring Survey
Russia	2007 Russian Longitudinal Monitoring Survey
Russia	2008 Russian Longitudinal Monitoring Survey
Russia	2009 Russian Longitudinal Monitoring Survey
Russia	2010 Russian Longitudinal Monitoring Survey
Russia	2011 Russian Longitudinal Monitoring Survey
Russia	2012 Russian Longitudinal Monitoring Survey
Russia	2013 Russian Longitudinal Monitoring Survey
Russia	2014 Russian Longitudinal Monitoring Survey
Russia	2015 Russian Longitudinal Monitoring Survey
Russia	2016 Russian Longitudinal Monitoring Survey
Rwanda	2016 Demographic and Health Survey
Rwanda	2018 Labor Force Survey
Rwanda	2019 Labor Force Survey
Rwanda	2020 Labor Force Survey
Saudi Arabia	2013 Household Expenditure and Income Survey
Solomon Islands	2005 Household Income and Expenditure Survey
Solomon Islands	2013 Household Income and Expenditure Survey
Sierra Leone	2003 Sierra Leone Integrated Household Survey
Sierra Leone	2014 Labor Force Survey
Sierra Leone	2018 Sierra Leone Integrated Household Survey
El Salvador	2000 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2001 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2002 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2003 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2004 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2005 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2006 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2007 Encuesta de Hogares de Propósitos Múltiples

El Salvador	2008 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2009 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2012 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2014 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2015 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2016 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2017 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2018 Encuesta de Hogares de Propósitos Múltiples
El Salvador	2019 Encuesta de Hogares de Propósitos Múltiples
Somalia	2013 Somalia Household Survey
São Tomé and Principe	200 National Survey on Household Living Conditions
Slovenia	2002 Household Budget Survey
Slovenia	2003 Household Budget Survey
Slovenia	2004 Household Budget Survey
Eswatini	2009 Household Income and Expenditure Survey
Eswatini	2016 Household Income and Expenditure Survey
Chad	2011 Enquête sur la Consommation des Ménages et le Secteur Informel au Tchad
Chad	2018 Enquête sur la Consommation des Ménages et le Secteur Informel au Tchad
Thailand	2010 Household Socio-Economic Survey
Thailand	2011 Household Socio-Economic Survey
Tajikistan	2009 Living Standards Measurement Survey
Tajikistan	2013 Jobs, Migration, Skills, and Consumption Survey
Timor-Leste	2010 Labor Force Survey
Trinidad and Tobago	2000 Integrated Public Use Microdata Series
Trinidad and Tobago	2011 Integrated Public Use Microdata Series
Tunisia	2000 National Survey on Population and Employment
Tunisia	2001 National Survey on Population and Employment
Tunisia	2005 National Survey on Population and Employment
Tunisia	2010 National Survey on Population and Employment
Tunisia	2011 National Survey on Population and Employment
Tunisia	2015 National Survey on Population and Employment
Turkey	2002 Household Income and Consumption Expenditure Survey
Turkey	2003 Household Income and Consumption Expenditure Survey
Turkey	2004 Household Income and Consumption Expenditure Survey
Turkey	2005 Household Income and Consumption Expenditure Survey
Turkey	2009 Labor Force Survey
Turkey	2010 Labor Force Survey
Turkey	2015 Labor Force Survey

Turkey	2016 Labor Force Survey
Turkey	2019 Labor Force Survey
Tanzania	2000 Integrated Labor Force Survey
Tanzania	2006 Integrated Labor Force Survey
Tanzania	2009 National Panel Survey
Tanzania	2011 Household Budget Survey
Tanzania	2012 National Panel Survey
Uganda	2005 Uganda National Household Survey
Uganda	2012 Uganda National Household Survey
Uganda	2016 Uganda National Household Survey
Ukraine	2002 Household Living Conditions Survey
Ukraine	2004 Household Living Conditions Survey
Ukraine	2005 Household Living Conditions Survey
Ukraine	2006 Household Living Conditions Survey
Ukraine	2007 Household Living Conditions Survey
Ukraine	2008 Household Living Conditions Survey
Ukraine	2009 Household Living Conditions Survey
Ukraine	2010 Household Living Conditions Survey
Ukraine	2011 Household Living Conditions Survey
Ukraine	2012 Household Living Conditions Survey
Ukraine	2013 Household Living Conditions Survey
Uruguay	2000 Encuesta Continua de Hogares
Uruguay	2001 Encuesta Continua de Hogares
Uruguay	2002 Encuesta Continua de Hogares
Uruguay	2003 Encuesta Continua de Hogares
Uruguay	2004 Encuesta Continua de Hogares
Uruguay	2005 Encuesta Continua de Hogares
Uruguay	2006 Encuesta Continua de Hogares
Uruguay	2007 Encuesta Continua de Hogares
Uruguay	2008 Encuesta Continua de Hogares
Uruguay	2009 Encuesta Continua de Hogares
Uruguay	2010 Encuesta Continua de Hogares
Uruguay	2011 Encuesta Continua de Hogares
Uruguay	2012 Encuesta Continua de Hogares
Uruguay	2014 Encuesta Continua de Hogares
Uruguay	2015 Encuesta Continua de Hogares
Uruguay	2016 Encuesta Continua de Hogares
Uruguay	2018 Encuesta Continua de Hogares
Uruguay	2019 Encuesta Continua de Hogares
Uruguay	2020 Encuesta Continua de Hogares
United States	2010 Current Population Survey
United States	2011 Current Population Survey

United States	2012 Current Population Survey
United States	2013 Current Population Survey
United States	2014 Current Population Survey
United States	2015 Current Population Survey
United States	2016 Current Population Survey
United States	2017 Current Population Survey
United States	2018 Current Population Survey
Uzbekistan	2003 Household Budget Survey
Venezuela	2000 Encuesta de Hogares de Propósitos Múltiples
Venezuela	2001 Encuesta de Hogares de Propósitos Múltiples
Venezuela	2002 Encuesta de Hogares de Propósitos Múltiples
Venezuela	2003 Encuesta de Hogares de Propósitos Múltiples
Venezuela	2004 Encuesta de Hogares de Propósitos Múltiples
Venezuela	2005 Encuesta de Hogares de Propósitos Múltiples
Venezuela	2006 Encuesta de Hogares de Propósitos Múltiples
Vietnam	2007 Labor Force Survey
Vietnam	2008 Labor Force Survey
Vietnam	2009 Labor Force Survey
Vietnam	2010 Labor Force Survey
Vietnam	2012 Labor Force Survey
Vietnam	2014 Labor Force Survey
Vietnam	2016 Labor Force Survey
Vietnam	2019 Labor Force Survey
West Bank and Gaza	2000 Labor Force Survey
West Bank and Gaza	2001 Labor Force Survey
West Bank and Gaza	2002 Labor Force Survey
West Bank and Gaza	2003 Labor Force Survey
West Bank and Gaza	2007 Labor Force Survey
West Bank and Gaza	2017 Labor Force Survey
Kosovo	2008 Labor Force Survey
Kosovo	2012 Labor Force Survey
Kosovo	2013 Labor Force Survey
Kosovo	2014 Labor Force Survey
South Africa	2008 Quarterly Labor Force Survey
South Africa	2009 Quarterly Labor Force Survey
South Africa	2010 Quarterly Labor Force Survey
South Africa	2012 Quarterly Labor Force Survey

South Africa	2013 Quarterly Labor Force Survey
South Africa	2015 Quarterly Labor Force Survey
South Africa	2016 Quarterly Labor Force Survey
South Africa	2018 Quarterly Labor Force Survey
South Africa	2019 Quarterly Labor Force Survey
South Africa	2020 Quarterly Labor Force Survey
Zambia	2002 Living Conditions Monitoring Survey
Zambia	2008 Labor Force Survey
Zambia	2010 Living Conditions Monitoring Survey
Zambia	2012 Labor Force Survey
Zambia	2014 Labor Force Survey
Zimbabwe	2001 Income, Consumption and Expenditure Survey
Zimbabwe	2007 Income, Consumption and Expenditure Survey
Zimbabwe	2011 Income, Consumption and Expenditure Survey
Zimbabwe	2017 Income, Consumption and Expenditure Survey

A5: Transparency and Replicability

The data cleaning process and the construction of indicators is done using STATA. All Stata code used in cleaning and estimation for versions 1.0 and 1.1, 2.0 and 3.0 have been archived on GitHub here. The code has successfully undergone the process to become part of the World Bank's Reproducibility Research Repository and has been approved (details forthcoming).

The data used in WWBI are not publicly available as of June 2020.

9. References

Christofides, L. C. and M. Michael. 2013. “Exploring the public-private sector wage gap in European countries.” *IZA Journal of European Labor Studies* 2 (15): 1-53

de Castro, F., M. Salto, and H. Steiner. 2013. “The gap between public and private wages: new evidence for the EU.” European Comission European Economy Economic Papers No. 508. Brussels, Belgium: EC

Dong Y. and C.-Y. J. Peng. 2013. Principled missing data methods for researchers. *Springerplus* 2 (222):1-27

European Commission. 2014. Government wages and labour market outcomes. European Commission Directorate-General for Economic and Financial Affairs. Brussels, Belgium: EC

International Labor Organization. 2021. ILOSTAT Glossary: Public sector employment. Genève, Switzerland: ILO

International Monetary Fund. 2014. Government Finance Statistics Manual. Washington, DC: IMF

International Monetary Fund. 2016. Managing Government Compensation and Employment—Institutions, Policies, and Reform Challenges. Washington, DC: IMF

Gindling, T. H., Z. Hasnain, D. Newhouse, and R. Shi. 2020. Are public sector workers in developing countries overpaid? Evidence from a new global dataset. *World Development* 126 (2020): 1-16

Giordano, R., M. C. Pereira, D. Depalo, B. Eugène, E. Papapetrou, J. J. Perez, L. Reiss, and M. Roter. 2015. The Public Sector Pay Gap in a Selection of Euro Area Countries in the Pre-crisis Period. *Review of Public Economics*, 214 (3): 11-34

Jakobsen, J.C., C. Gluud, J. Wetterslev and P. Winkel. 2017. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology* 17 (162): 1-10

UNSCD, 2007, International Standard Industrial Classification of All Economic Activities Revision 4, Series M: Miscellaneous Statistical Papers, No. 4 Rev. 4, New York: United Nations. ST/ESA/STAT/SER.M/4/REV.4