

# The Gender Disaggregated Labor Database

## *Abstract*

The Global Disaggregated Labor Database will fill an important information gap in global gender statistics by providing detailed accounts on employment levels, wages, labor income, and employment status at detailed economic activity. Until now, only data in broad sectors is available for cross-country comparisons. This effort is intended to have a direct application in applications involving computable general equilibrium (CGE) modeling of the impacts of trade policy changes in developing countries. However, other types of modeling frameworks including macroeconomic modeling can benefit from such database.

## *Extended Abstract*

### 1 Motivation

**Gender is integral to the process of development.** Gender development outcomes are both strong determinants and co-dependent on a variety of development goals. At the household level, for instance, gender gaps in educational attainment are a strong determinant of households' allocation of labor and the roles played by family members in providing care; it is well-established that mothers' educational attainment is a better predictor of the educational achievement of household's offspring; largely explained by the fact that women, having different spending patterns, tend to allocate a larger share of expenditure on health, education, and well-being of their children. On aggregate too, gender is strongly interlinked with economic performance. Labor participation of women can play an important role in raising the rate of potential output, especially in countries where women's' labor force participation has been historically low. Similarly, economic activities that rely on women work force, such as manufacturing of wearing apparel and textiles, can shape the comparative advantage across nations, especially when opened and integrated with global value chains.

**As a result, the gender dimension is crucial to the formulation of sound economic policy.** It is well-established that a careful examination of the impact of economic policy should consider the gender dimension to address the multitude of impacts across different segments of the population– and that this examination should extend beyond the microeconomic approach. Macroeconomics, including trade policy, is increasingly considering gender a fundamental aspect for the design of economic policy. Not only macroeconomic policy can have sizable and long-lasting effects on gender-related outcomes; but also, existing gender inequities can influence the effectiveness of macro-economic policy. Typically, macroeconomists are interested in the effects on labor force participation, financial inclusion, trade diversification, firm performance, intra-household choices, and public investment.

**As of now, important data gaps on gender statistics on labor conditions still remain.** Particularly, comparable data across countries on detailed sector of employment is scarce. To the best of our knowledge, only employment at an aggregate level is published at the broader economic activity (agriculture, industry, services)<sup>1</sup>. The World Economic Forum's Global Gender Gap Report<sup>2</sup>, for instance, present aggregated indicators for earning gaps only for skilled workers and lacks sufficient detail on women's sector of employment, labor volumes, and earning by detailed economic activity. Broad aggregations in global gender statistics are insufficient for a careful examination of the links between international economic policy, gender and poverty.

## 2 Development Objective

**This project's development objective is to provide a global public good in the form of a documented database that complements statistics on employment and labor incomes disaggregated by gender and detailed economic activity.** This database aims to calculate the skill intensity within detailed sectors for global sample of countries. The collection of data will use the existing collections of household surveys available within the World Bank. Data on skill-intensity by economic activity and gender will complement the Global Trade Analysis Project (GTAP) database. This effort is intended to have a direct application in applications involving computable general equilibrium (CGE) modeling of the impacts of trade policy changes in developing countries. However, other types of modeling frameworks including macroeconomic modeling can benefit from such database.

## 3 Methodology

**The construction of this database requires the use of household surveys to collect statistics on employment, labor income, detailed sector of employment, gender, and skill level.** The construction of the Gender Disaggregated Labor Database is not intended to be a brand-new harmonization; rather, it is built on top and will be integrated with the ongoing household survey harmonization efforts supported by the World Bank. As a result, statistics and code used for the construction of the Gender Disaggregated Labor Database can immediately be integrated in the infrastructure that informs World Bank's flagships, research, and technical reports. The technical steps are summarized in the Table below, while a more detailed description of the project's steps can be found in Section 6.

---

<sup>1</sup> This data was produced by the International Labor Organization (ILO). It is available at the World Bank Gender Data Portal.

<sup>2</sup> <http://reports.weforum.org/global-gender-gap-report-2016/>

Table 1 List of steps involved in the construction of the Gender Disaggregated Labor Database

- a. Recollect meta-data on national systems of classification for economic activity, educational attainment, and occupation classifications. Main source: United Nations, International Labor Organization
- b. Identify and prioritize household surveys subject to be further harmonized within the World Bank catalogs<sup>3</sup>
  - a. Prioritize based on comparability, size of GDP, population, coverage
- c. Harmonization of detailed sector of economic activities
  - a. Methodology has been applied to subset of countries: Colombia, Egypt, Indonesia, Cambodia, Lao, Sri-Lanka, Lesotho, Malaysia, Thailand, and Vietnam
  - b. Building concordance tables
  - c. Defining rules for splitting volumes
  - d. Documentation
- d. Develop cross-tabulations and construction of public database
- e. Dissemination

Source: Authors' elaboration

## 4 The Gender Disaggregated Labor Database

### 4.1 Internal Output within the World Bank Group

**The internal output consists of adding harmonization variables and code to the existing harmonization collections hosted within the World Bank.** The new harmonization of the industry and occupation variables will be added to the existing harmonization code for the selected surveys. New developed code and variables can be accessed by all World Bank datalibweb users. This new information hence will have two forms. First, the variables `industry_original` and `occupation_original` will be created to provide a common variable name across surveys for retrieving the variable that identifies the sector of employment of the first and second job of every employed individual. This `industry_original` and `occupation_original` variables will have information on the classification system used in each survey. Second, the variable `industry_gtap` will be added to the standard set of harmonized variables. This variable is comparable to the disaggregation of activities in the GTAP database.

---

<sup>3</sup> These harmonization catalogs are maintained by the Poverty and Equity Global Practice (POV) through Datalibweb, which is a data system specifically designed to enable World Bank users to access the most up to date version of non-harmonized (original/raw) and harmonized datasets of different collections across World Bank Global Practices.

## 4.2 External output

**The main external output is a vetted aggregated database that can be accessed through the World Bank data portals.** An aggregated version of this database can be used for public dissemination. This version has to be vetted by World Bank country economists considering the adequacy of the calculations and providing clearance to provide aggregate statistics from household survey data. A preliminary version of this data is expected to be released in November.

The external database will have the following variables, and several indicators can be calculated for public dissemination:

### **Section A: Survey/Census Identifier**

- Country 3-letter ISO Code: World Bank
- Country 3-letter ISO Code: United Nations
- Country Numeric Code: United Nations
- Country Code: GTAP
- Country Name
- Survey or Census Code:
- Survey or Census Name:
- Survey year:

### **Section B: Labor Volumes**

- Minimum Age for Labor Questionnaire:
- Labor Volumes for the Following Groups, by Gender:
  - Population Below Minimum Age for Labor Questionnaire:
  - Population Covered by Labor Questionnaire
  - Population Not in the Labor Force
  - Unemployed
  - Labor Incomes for the Following Sub-Groups:
    - Employed, Wage Earners
    - Employed, Self-Employed
    - Employed, Employers

### **Section C: Labor Income**

- For employed wage earners, labor Incomes by:
  - Gender
  - Skill Level Using Four Levels of Educational Attainment
    - No Education or Less Than Primary Completed
    - Primary Completed but Less Than Secondary Completed
    - Secondary Completed but Less Than Tertiary Completed
    - Tertiary Completed
  - Skill Level Using Occupation Classification
  - Activity of employment: using the GTAP version 10 classification of sectors

**Additionally, this database will be showcased in an external World Bank website.** A simple website with standard dashboard and live dashboards can be set up to provide live statistics on the database. Tableau Dashboards are the preferred standard for performing simulations. In addition to the gender data website, the data

should be disseminated via the institutional data outlets, such as the World Development Indicators, or the Gender Data Portal.

**The external output includes a paper/background note that support the database.** This background paper will document the construction of the database, including methods, household surveys used, assumptions. Additionally, it should provide provides stylized facts about its general use and the integration with the GTAP database. Lastly, the background paper will include areas of further development.

## 5 ANNEX: Risk of the project

### 5.1 Risk related to the process of harmonization

**There is an important risk associated with the nature of developing harmonization code for household surveys,** especially if the outputs have serious limitations in terms of documentation. In efforts like this, it is natural that errors occur, without procedures to verify and without proper documentation, there is a high risk of reputational damage to the database. These types of risk can be mitigated by developing regular data checks on summary statistics and compare findings across countries and against other sources (i.e. national accounts). Additionally, there should be a process of vetting this data with country economists.

### 5.2 Risk associated with ownership of harmonization standards

**There is a risk of ownership.** The risks of ownership of household survey harmonization code can delay access to the data, fragment cooperation within the World Bank. These risks can be mitigated by: 1) Explaining that the project aims to expand current harmonization efforts, especially that of the "industry" and "occupation" variables. This harmonization project can be useful to GMD, I2D2, or regional databanks because it will collect information about the National Classification System for each country and survey. Additionally, the project will prioritize surveys that are common to existing harmonization efforts. For instance, Mexico's ENIGH, Vietnam's VHLSS, China's CHIPS, and Europe's EU-SILC.

### 5.3 Risk associated with using alternative surveys

**There is a risk associated with not using labor force surveys.** The majority of databases within the World Bank harmonization systems deal with the measurement of welfare and poverty. There are risks associated with working on top of a databases that are not ideal for the measurement of labor statistics. It was suggested to collect labor force surveys and compare estimates. This can be done for a limited set of countries. It is important to acknowledge that income/ budget surveys and labor force surveys can have very different aggregate labor statistics for the same year and even in countries with good statistical capacity.

## 6 ANNEX: Constructing a global wage bill database

### 6.1 Collect meta-data on national systems of classification for economic activity

- Compile a cross-country meta-database with information about national classification systems for (a) industry, (b) occupation, (c) education.
- This meta-data was collected by the [United Nations Statistical Division](#) and by the International Labour Organization.
  - UNSD has carried out surveys on the actual use of classification in countries to monitor the implementation of international statistical classifications and to obtain an overview of (a) adjustment that countries make to these classifications for national purposes, (b) the applications of these classifications and (c) the revision plans that countries have for such classifications.
  - UNSD last survey was started in 2012, but countries have continued to submit updated. A next round of this survey was planned for 2018 but has not yet started.
  - UNSD's national classifications database contains information on 909 classifications from 159 countries (475 current classifications and 434 previously used classifications). These represent primarily classifications of activities, products, occupations and expenditures, but also include health and education classifications for some countries.
  - The UN website was discontinued but can be accessed through the WebArchive at:

### 6.2 Identify and prioritize household surveys subject to be further harmonized within the World Bank catalogs

Prioritize according to:

1. Ease of harmonization and quality of data
2. The relative importance of their exports in global and regional trade
3. The size of GDP within the World Bank income classification
4. Size of Population within the World Bank regional classification

### 6.3 Harmonization of detailed sector of economic activity

The procedure involves 3 steps. First, we constructed concordance tables between different systems of classification. Second, we defined a set of rules for moving monetary values from broader to more detailed classification systems. The final step involves the calculation of the skill-intensity within the manufacturing sector at the GTAP level of disaggregation.

6.3.1 Step 1: Building concordance tables

We built concordance tables between 3 classification systems of economic activity:

- i) International Standard of Industrial Classification Revisions 4, 3.1, and 3 (ISIC Rev. 3)
- ii) Global Trade Analysis Project (GTAP)
- iii) Harmonized System (HS)

A sub-sample of countries (Vietnam, Thailand, Sri Lanka, Cambodia, Lesotho, Indonesia, Colombia, Egypt) were selected for a feasibility pre-exercise:

- i) There exists a previous harmonization of *industry, education, and occupation* variables
  - a. We relied in data and harmonization code from the International Income Distribution Database (I2D2) from the World Bank
- ii) The national industry classification is based on a recent version of ISIC (Rev 3, 3.1, or 4)
  - a. This information can be obtained from UN questionnaires as seen in Table 1 below.

**Table 2: Egypt – National Classification of Economic Activity. Sample questionnaire**

Relationship to international standards		
3 (a)	Is this classification based on (or linked to) an international standard classification? If yes, please describe.	<ul style="list-style-type: none"> <li>• At 4 digits.</li> </ul>
3 (b)	Is the classification structure identical to the international standard or, if not, how does it differ? (e.g. have additional levels been added to the international standard or have changes been made within the level of the international structure, such as aggregations or additional breakdowns)	<ul style="list-style-type: none"> <li>• Yes, Now Egypt uses the national 2007</li> <li>• <b>According to the international classification (ISIC Rev.4.)</b></li> </ul>

source: Table based on questionnaire prepared by the United Nations Statistical Division. Available at <https://unstats.un.org/unsd/cr/ctryreg/ctrydetail.asp?id=1103>. Accessed July 28, 2017

This data is not available anymore on the UN Stats website. Instead use the following link:

6.3.2 Step 2: Defining rules for splitting volumes

Translating monetary values between systems of classification involves the use of researcher judgement. This inevitably occurs because there is not a strict 1-to-1 relationship between classification codes; rather, a one-to-many concordance is found frequently. Assigning monetary values from a broader to a more detailed classification researchers must have assignation rules. In our context, we constructed weights based on exports values at the more detailed Harmonized System classification.

### 6.3.3 Step 3: Calculating the skill intensity

Relying on existing harmonization code within the World Bank systems (i.e. I2D2 harmonization code), we created a newly harmonized *industry* variable compatible with the GTAP database and the LINKAGE model. Skill intensity based on both i) occupation and ii) education were calculated for each country. The household survey sampling weights were scaled-up to reflect the total population in the base year of the GTAP database (2011).

## 7 ANNEX: Frequently Asked Questions

### 7.1 What is I2D2?

The **International Income Distribution Database** (I2D2) is a global harmonization effort hosted by the World Bank. The I2D2 evolved from the need of presenting a global picture on key development aspects related with global trends on inequality, education, labor, and employment; mostly to inform World Development Report. Early efforts to construct a World Development Report 2006: Equity and Development.

### 7.2 What is GMD?

The **Global Micro Database** (GMD) is a similar harmonization effort. Launched, it is now the official database for the measurement of the World Bank Twin Goals of Poverty and Shared Prosperity.

### 7.3 Who can request access to I2D2, GMD or similar databanks within the World Bank?

Access to the harmonized version of I2D2, GMD, or other databanks projects through datalibweb. A Stata program that manages access to the different collections of household surveys within the World Bank.

For I2D2, the policy is that all staff with a World Bank Group e-mail address (for instance: @worldbank.org or @ifc.org) can request access to the harmonized version of the micro-data. A written-request must be sent to David Newhouse, Senior Economist adding the research purpose and the names for all those requesting accessing the harmonized version of the data.

There are few members involved in the I2D2 development that have access to the original databases and the harmonization code, which is hosted within the World Bank internal servers.

### 7.4 Why the use of I2D2 or GMD?

While access to the harmonization code is the main limitation, there is a large overlap in coverage and harmonization variables between I2D2 and GMD. For the case

*ANNEX: Frequently Asked Questions*

of Latin America and the Caribbean, the I2D2 and GMD are derived from the SEDLAC harmonization. In other cases, they are independent.