OUTLINE

1. Motivation for constructing a new database
2. Methodology
3. Deliverables
Motivation
Motivation

Need for a (re)examination of the distributional consequences of international trade beyond macroeconomic indicators

Nevertheless, detailed comparable data for global policy analysis is scarce, particularly for impacts by gender

Objective: Create a comparable micro-based dataset that can inform macro (CGE) models, providing detailed accounts on employment levels by gender, occupation, wage, employment status at a finer level in economic activity.
Future applications within the World Bank

a) **Processed “ready-made” statistics,**
   To be used in the upcoming World Bank (2020) *Trade and Gender Report*
   Cross-country stats such as women's labor intensity and gender gap by sector

b) **Statistics in flexible format** to inform other databases.
   To be used as an input to the Global Trade Analysis Project (GTAP) Database
   Disaggregate labor accounts by gender.

c) **Added variables to World Bank micro-data surveys**
   To be used in simulations or generation of statistics
   i.e. Impact of trade policy on WB client countries, AfCFTA, CP-TPP, USMCA
2 Methodology
Starting point

• The World Bank has harmonized data for 138 countries with broad industry (3 sectors or 10 sectors) and occupation variables that account for 90% of population and GDP
Starting point: Global

Female labor intensity:
The share of female employment in an economic activity divided by the share of female employment in the country

Source: World Bank staff estimates using the International Income Distribution Database
Harmonization Process for the Gender Disaggregated Labor Database

Creates a finer level of disaggregation for “industry” and “occupation” harmonized to international standards.

Global collection of Household Surveys

Recover original industry and occupation variables

Is concordance possible?

Yes

No

Document reason:
- a) # of digits is not enough; or
- b) Concordance missing
- c) Other

Meta-data: Collect national classification systems for Industry and Occupation

ISIC & GTAP (2-digit)

ISCO 08
From micro-based statistics to national aggregates matching national accounts (GTAP)

Harmonized micro data

Cleaning Process
- Within-country outlier identification & treatment

Cross-country outlier identification & treatment

Initial wages & employment by activity, gender and skill level

Validation Adjustment of labor value added & employment to ILO, WB, and GTAP

Estimated Aggregated Dataset

Working at micro level

Working at aggregate level
Building a Disaggregated Labor Database, limitations

• On the microeconomic side, it requires a re-harmonization of industry and occupation data relying on a sub-sample of the World Bank collection of household surveys.

• The strategy has been to find initial documentation for a cross-section of countries.

• Finding disaggregated data at industry level in household surveys is not easy, because of:
  • Focus. The World Bank efforts to harmonize household survey data are concentrated on collecting data for monitoring “global poverty”. This has created a gap in survey coverage for a. data on high-income countries, and b. an emphasis in poverty, rather than in the production of labor statistics.
  • Survey design. Either when the classification system present in the household survey cannot be further disaggregated. Due to sampling, most surveys are not representative at the highest disaggregation level.
  • Data quality: Even if data can be disaggregated, there are some internal inconsistencies, inherent to data collection in difficult environments and with teams with low-statistical capacity.
  • Inconsistent meta-data: Comparable international information about national industry and occupation classifications doesn’t exist and country information is not always easily available.
Building a Disaggregated Labor Database, progress on meta-data collection

- 219 countries-surveys were revised. When possible, information was collected about national industry and occupation classification, including number of digits in the survey.
Building a Disaggregated Labor Database, progress on performing conversion

- **a. ISIC Rev 4 and GTAP v10 codes**

<table>
<thead>
<tr>
<th>Industry</th>
<th>East Asia &amp; Pacific</th>
<th>Europe &amp; Central Asia</th>
<th>Latin America &amp; Caribbean</th>
<th>Middle East &amp; North Africa</th>
<th>North America</th>
<th>South Asia</th>
<th>Sub-Saharan Africa</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harmonized in individual-level format</td>
<td>11</td>
<td>12</td>
<td>15</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>22</td>
<td>76</td>
</tr>
<tr>
<td>Harmonized from Luxembourg Income Study</td>
<td></td>
<td>13</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Not harmonized, concordance missing</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Not harmonized, not enough digits</td>
<td>4</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Not harmonized, raw Data before 2000</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Microdata is not available in WB domain</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Microdata do not exist (territories/small states)</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>Microdata do not contain industry variable</td>
<td>3</td>
<td>2</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>38</strong></td>
<td><strong>59</strong></td>
<td><strong>42</strong></td>
<td><strong>21</strong></td>
<td><strong>3</strong></td>
<td><strong>8</strong></td>
<td><strong>48</strong></td>
<td><strong>219</strong></td>
</tr>
</tbody>
</table>
Building a Disaggregated Labor Database, progress on performing conversion

- and/or b. ISCO 08 codes have been processed for:

<table>
<thead>
<tr>
<th>Industry</th>
<th>East Asia &amp; Pacific</th>
<th>Europe &amp; Central Asia</th>
<th>Latin America &amp; Caribbean</th>
<th>Middle East &amp; North Africa</th>
<th>North America</th>
<th>South Asia</th>
<th>Sub-Saharan Africa</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harmonized in individual-level format</td>
<td>11</td>
<td>7</td>
<td>13</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>19</td>
<td>63</td>
</tr>
<tr>
<td>Harmonized from Luxembourg Income Study</td>
<td></td>
<td>13</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Not harmonized, concordance missing</td>
<td>4</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td></td>
<td>1</td>
<td>14</td>
<td>37</td>
</tr>
<tr>
<td>Not harmonized, not enough digits</td>
<td>3</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td></td>
<td>3</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Not harmonized, raw Data before 2000</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Microdata is not available in WB domain</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>8</td>
<td></td>
<td>2</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Microdata do not exist (territories/small states)</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td></td>
<td>1</td>
<td></td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Microdata do not contain industry variable</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>38</strong></td>
<td><strong>59</strong></td>
<td><strong>42</strong></td>
<td><strong>21</strong></td>
<td><strong>3</strong></td>
<td><strong>8</strong></td>
<td><strong>48</strong></td>
<td><strong>219</strong></td>
</tr>
</tbody>
</table>
An application example for finer level industry

- Two prominent options for diversification in Africa are a. Vegetables and Fruits and b. Textiles
- Vegetables and fruits have higher value added than traditional agricultural products such as crops, coffee, cocoa
- Textiles, on the other hand, have been the base of manufacturing expansion in East Asia
- Due to high intensity of female labor and gender wage premia, both sectors can have different implications in job creation, by gender
3 Deliverables
Building a Disaggregated Labor Database, what has been done

The initial deliverables are:

1. Harmonized micro data (household survey)
   • with labor module including employment status, labor status, demographic information, wage and income, ISCO Rev 4 (2-4 digit) and GTAP10, ISCO 08 (2-4 digit)
   • Only open for the users who have the access to the WB household survey data

2. Harmonization Tool (in Stata)
   • Allows user to harmonize a new household survey between national and international classifications. This is critical for future scaling-up this project.

3. Global statistics on wages and employment by economic activity, skill level and gender
   • This global statistics match GTAP v10 value-added data

4. Meta-data on national classification systems

5. Document and Visualization – to be released in March 2020*
   • World Bank Website
   • World Bank Open Data Bank
   • World Bank Data Catalog
Dataset: Harmonized micro data for internal use only

- The GDLD database complements global and regional harmonization efforts within the World Bank,
  - Particularly, economic activities and occupation variables were re-harmonized to match international classifications systems (ISIC Rev 4/ISCO 08)
  - Additionally, a proxy variable of “wage per month” was created
  - Because it is based on record-level data, the dissemination of the dataset is restricted to only within the WBG. Nevertheless, the underlying code can be shared without restriction
  - At the global level, the harmonized micro-database accounts for 71% of GDP and 78% of population
**Dataset: Regional coverage of harmonized micro data**

<table>
<thead>
<tr>
<th>Regions</th>
<th>Population</th>
<th>GDP</th>
<th>Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>millions</td>
<td>US$Billions</td>
<td></td>
</tr>
<tr>
<td>East Asia and the Pacific</td>
<td>2,329</td>
<td>26,058</td>
<td>11</td>
</tr>
<tr>
<td>Europe and Central Asia</td>
<td>921</td>
<td>23,161</td>
<td>25</td>
</tr>
<tr>
<td>Latin America and the Caribbean</td>
<td>610</td>
<td>5,215</td>
<td>17</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>449</td>
<td>3,674</td>
<td>7</td>
</tr>
<tr>
<td>North America</td>
<td>364</td>
<td>22,283</td>
<td>1</td>
</tr>
<tr>
<td>South Asia</td>
<td>1,814</td>
<td>3,454</td>
<td>8</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>1,087</td>
<td>1,706</td>
<td>22</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>7,574</strong></td>
<td><strong>85,550</strong></td>
<td><strong>91</strong></td>
</tr>
</tbody>
</table>

**Notes:** Includes surveys from World Bank repositories and aggregates from Luxembourg Income Study. Based on World population and GDP in 2018.
Tool: Stata ado-files (re)-run the harmonization

• **gdld_class**
  • Provides a framework to harmonize national industrial or occupational classifications to their ISIC Rev 4, GTAP v10 or ISCO-08 counterpart

• formed by a set of STATA programs that runs on top of individual-level micro data

• It helps users to replicate the harmonization done in this project and run their own harmonization based on the local classification information

• It is a critical piece in scaling this project and sharing it with units inside and outside the World Bank
Dataset: Estimated wages and employment volume (for CGE modeling)

- Estimated statistics based on the harmonized household survey dataset
  - It provides the estimated wages and employment volumes disaggregated by economic activity, skill level and gender
    - Particularly useful for the GTAP 10 database with 65 sectors and 141 regions
  - This database has been adjusted to match macro statistics on labor force and wage bill (available in World Development Indicators and ILOSTAT)
- It includes WB and GTAP regional mappings
Metadata: national classification systems for economic activity and occupation

• Metadata collected throughout this project.

• It includes information for economic activity and occupation variables present in World Bank household surveys

• The metadata was collected from extensive search, including the national statistical offices, survey documents, external websites, and other sources

• It helps to identify national classifications used by the surveys, in order to create the harmonization process from local classification to international classifications
Platform: GDLD Website

• GDLD website provides a glance for this project, to download necessary data and files, visualize the database for global comparison, and link to applications such as Trade and Gender Report and AfCFTA project.
  • The internal review link is http://datatopicsqa.worldbank.org/gess/
  • The final public URL would be http://datatopics.worldbank.org/gdld/
Platform: Data Bank Page

- It is an interactive platform allowing users to select, preview and download necessary summary data. GDLD Data Bank page is 4-dimension for country, sector, series and time. In current stage, it is in internal review status.
Platform: Data Catalog Page

- The Data Catalog page (https://datacatalog.worldbank.org/ddhworkbench/workbench-datasets) is the place to save all materials, which will be shared to public after the review.