

# Modeling the poverty and distribution effects of human capital formation across countries:

Using a new database on labor skills and wages

Marcio Cruz, Delfin S. Go, and Israel Osorio-Rodarte<sup>1</sup>

PRELIMINARY DRAFT – PLEASE DO NOT CITE OR QUOTE

Comments and suggestions are welcome

The World Bank

April 17, 2017

## Abstract

Modeling the poverty and distribution consequences of human capital formation requires consistent treatment of labor and its skill composition between the economic model and the household data. This paper describes the innovations made to Global Income Distribution Dynamics (GIDD), a macro-micro simulation framework at the World Bank to capture the global poverty and distribution effects of long-term demographic and education trends across countries. First, it introduces a new database about the skill composition of labor based on education level as well as demographic characteristics of individuals, derived from harmonized household surveys of 104 countries. It is constructed to provide a consistent definition of labor skills, wages and skill premia to disaggregate the value added of labor in social accounting matrices from national accounts. In so doing, it summarized useful stylized facts for economic modeling and provided a means to access the resulting data. Second, it explains how the new data is incorporated into a global microsimulation model like the GIDD, possibly as a template for parallel economic modeling. And third, it summarizes two cases among the many applications of the framework already made – the first case deals with the global poverty and inequality implications of the coming wave of educated workers from the developing countries, and the second case examines the future threats of the antimicrobial resistance.

---

<sup>1</sup> All from Development Prospects Group, The World Bank. Emails are Cruz: [marciocruz@worldbank.org](mailto:marciocruz@worldbank.org); Go: [dgo@worldbank.org](mailto:dgo@worldbank.org); and Osorio-Rodarte: [iosoriorodarte@worldbank.org](mailto:iosoriorodarte@worldbank.org).

# 1 Introduction

The economic modeling of poverty and income distribution projecting forward-looking scenarios requires an integration of a macro model that examines the economy-wide effects of economic trends and shocks with microdata that traces the implications on households and individuals. Over long time horizons, changes in demographic structure (share of the working-age population, gender, etc.) and in the skill composition of labor (the effects of education, training, etc.) become important in the accumulation of human capital and its subsequent effects on growth, poverty, and income distribution. Even so, the integration of demographic and labor characteristics is largely difficult because of the dissimilar treatment of labor, skills, and wages between the economic framework at the macro level and the household information. This paper describes a recent effort to improve such integration through the creation of a new database about the skill composition of labor, its implications on wages, and its incorporation into the Global Income Distribution Dynamics (GIDD), a global macro-micro simulation model tool developed at the World Bank's Development Prospects Group.

The GIDD's extensive coverage of microdata from household surveys permits the ex-ante investigation of issues not easily tractable with other methods, especially across countries. For example, Ahmed et al. (2017) decompose the impact of demographic and education trends on global inequality into a component due to changes in average incomes between countries and a component due to a widening disparity of household incomes within countries. Ahmed and Cruz (2016a) also look at the effects of demographic change and several policies aiming to boost demographic dividends on economic growth and poverty reduction. Another recent study, Ahmed et al. (2016b), examines the significance of Sub-Saharan Africa's demographic dividend for its future growth and poverty. In Devarajan et al. (2015), the authors stress-test Africa's recent growth and poverty performance. Lakatos et al. (2016) examines the potential impact of China's slowdown on Africa's growth and poverty. Another application, Balistreti et al. (2016), looks at its prospects for deep integration in eastern and southern Africa and potential effects on poverty and shared prosperity. Moreover, the tool has been directed to analyze the potential economic and poverty effects of pandemics, such as the future threat of the antimicrobial resistance in World Bank (2016). It could also be used to assess the growth and distributional effects of global targets, such as those defined by the United Nations Sustainable Development Goals, or any global policies about trade, climate change, or migration affecting both high- and low-income countries alike.

As a tool that provides ex-ante analysis of the effects of global macro shocks at the heterogeneous household level across countries, the GIDD model is among the most sophisticated top-down method available. It links a large amount of household data for over 130 countries to a global computable general equilibrium (CGE) or any global macroeconomic model. The first version of GIDD was implemented and described in Bussolo, De Hoyos, & Medvedev (2010) and Bourguignon & Bussolo (2013). It extends the single-country macro-micro framework in Bourguignon and Savard (2008) to a global setting.

Figure 1: A general representation of GIDD 2.0

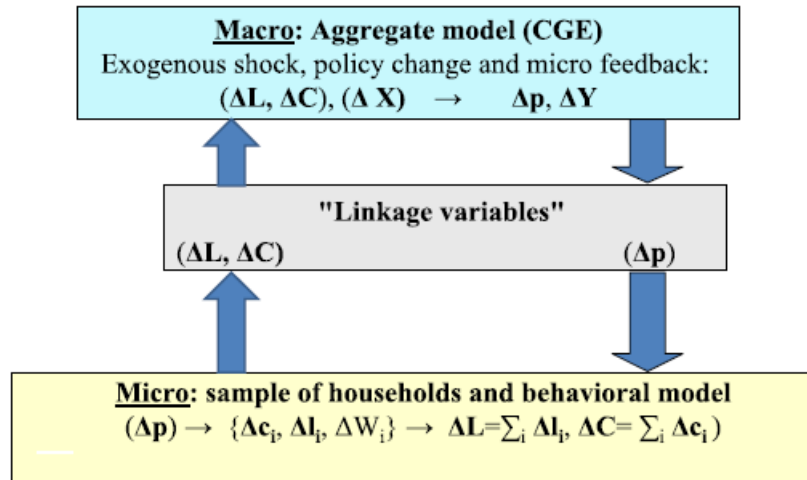


Figure 1 illustrates the broad representation of a macro-micro framework such as the GIDD model. At the bottom or household level of the macro-micro framework, GIDD follows the tradition of microsimulation works of Bourguignon et al. (2002, 2008), Bourguignon, Pereira de Silva (2003), and Davies (2009), which estimate the effects of policy change or shocks on households and income distribution. The top level of the framework is often a global computable general equilibrium (CGE) model, such as the LINKAGE (Van Der Mensbrugghe 2011), GLOBE (McDonald et al. 2007), or Dynamic GTAP model GDyn (Ianchovichina and Walmsley 2012). These global macro models provide the economy-wide implications of macroeconomic projections, exogenous shocks, or policy change on relative prices and factor rewards, including wages (i.e., the inputs that go into the GIDD). However, the GIDD could also be linked to single-country models or simple macroeconomic projections, and it could be used an independent tool whenever an analysis calls for the heterogeneous effects of shocks and policies on household or labor income.

The consistency and extent of information exchange in the linkage variables of the framework drives the quality of the analysis. In single-country applications such as the Philippines (Bourguignon and Savard 2008) or South Africa (Go et al. 2010), the availability of data enables an income-generation model with an elaborate occupational-choice behavioral model at the household level. That approach permits wages and employment probability of individuals to become the functions of diverse household characteristics, such as age, gender, level of schooling, years of work experience, sector of employment, urban or rural area, family size, head of household (or not), etc. In global models, a minimum set of information has to be uniform across the SAMs and household surveys of multiple countries. With the links defined, relative and factor prices, generated from the macro side are fed into the behavioral model at the household level, essentially regenerating the household survey(s) with a pattern of household employment, income, and consumption as well as relevant poverty and inequality measures that are consistent with the new prices and wages.

Updates of aggregates of labor and consumption from the micro side then become new inputs to the macro side in an iterative or dynamic recursive manner (see Bourguignon and Bussolo (2013) for more discussion).

Even so, a critical area in the consistency of the linkage variables remains the treatment and definition of labor, particularly the skill composition of labor. For example, global CGE modeling such as Tyer and Shi (2012) has employed estimates and projections of working-age populations by the United Nations to examine the effects of demographic change. Nevertheless, consistent definition of labor skills at both the top and bottom level is rare. Although this issue also affects single-country applications, it is more easily handled by selecting countries with the necessary data for resolution. In global applications, the underlying differences in the definition are harder to resolve; large discrepancies in the data and definition of the macro and micro level can impose limitations to the forward-looking scenarios.

A large part of the issue is the employment of different sources of data with significant dissimilarities in the definition of labor skills. Social accounting matrices (SAMs) used in global CGE modeling, such as those compiled for many countries by the Global Trade Analysis Project (GTAP) at Purdue University (Narayanan, et al., 2012), derives the value added of labor from national accounts. When only aggregate labor is needed, labor can be defined in efficiency units from the wage bill of labor. When skill composition is needed, the classification of skills often uses broad professional occupation such as managers, clerks, engineers, etc., following the definition and statistics of International Labour Organization (ILO). Information about the type of occupation is, however, not widely available in household surveys. Instead, the level of education is often provided for each household member. The level of schooling is also easier to forecast in an ex-ante analysis (given known enrollment rates and assumptions about their progress) than the various types of occupation. Also, GIDD version 1.0 does not exploit the individual characteristics available in the household surveys; each household essentially collapses into a representative individual (the head of household). Hence, it seems worthwhile to expand and reconcile the definition of labor according to the household characteristics available in the surveys, including gender, labor supply according to working-age composition, and the definition of skills based on schooling.

Using the level of schooling as a proxy for skill or quality of human capital is not new. It follows the practice of cross-country analysis of growth such as Barro (1998), Hall and Jones (1999), Klenow and Rodríguez-Clare (1997), among others. Despite some measurement issues (see Hendricks 2002), schooling has three advantages for global or country microsimulation: i) data about years of schooling has a greater availability and a much larger coverage than the occupation variable in household surveys across countries; ii) prospective educational assumptions for ex-ante analysis are less restrictive and more transparent than occupational ones; and, iii) the education variable has more direct associations with policy instruments and outcomes.

This paper aims to contribute to the literature on poverty and income distribution using microsimulation tools by describing a new database on skill composition of labor based on education and demographic characteristics of individuals, derived from harmonized household surveys of 104 countries. This data was constructed to provide a consistent definition of labor skills, wages and skill premia for economic modeling. It supplies useful and scarce information about labor income to disaggregate and improve the value added of labor in national accounts (such as the SAMs used in CGE modeling), making it more tractable for policy analysis. The database provides the wage premia of skilled labor versus unskilled labor, including the number of employees and average wage by sector of activities and level of school attainment in many countries. The new information helps to uncover significant stylized facts on global labor income distribution. As amplified below, it shows a strong and positive correlation between wages for both skilled and unskilled workers and GDP per capita with higher dispersion among lower and middle-income countries. Even so, a weak negative correlation appears to exist between skill wage premia and GDP per capita. Furthermore, we also observe a negative correlation between the share of unskilled workers in the wage bill and income per capita. Finally, we show a positive correlation between skill wage premia and Gini coefficient.

The remainder of the paper is organized as follows. Section 2 described the wage bill database by sector and level of skills. Section 3 explains how the new data is incorporated into a global microsimulation model like the GIDD. Section 4 summarizes few cases of the many applications of the framework already made – the first case deals with the global poverty and inequality implications of the coming wave of educated workers from the developing countries, the second case analyzes the effects of demographic change in Africa, and the third case examines the future threats of the antimicrobial resistance. Section 5 concludes.

## 2 A database about labor skill and wages

Although critical for improving our understanding of the relationship between poverty, global inequality and labor market dynamics, information about labor income by sector and level of educational attainment as well as workers' individual characteristics that allow comparison across countries is scarce. To fill this gap, we gather wage and other sources of income information from household surveys of numerous countries to build a new database about skill composition by sectors in GIDD 2.0, which we call the Global Income Distribution Dynamics Sectoral Wage Bill Database (GIDD-SWB). The data derives its information from the harmonized household surveys available in the International Income Distribution Database (I2D2) of the World Bank (see Montenegro and Hirn 2009) to permit key modeling innovations in the GIDD framework. These innovations include: i) new method of recalibrating surveys' sampling weights; ii) the distinction between different sources of monetary income; iii) the distinction of labor based on characteristics such as working age,

gender, and schooling; and iv) the use of country-specific household demand curves to account for changes in relative prices of food versus non-food consumption bundles.

The main contribution of the GIDD-SWB database lies in its global coverage and disaggregation by industry and workers' level of education and other characteristics. Also, wages are consistently reported in both US\$ of 2007 and 2005 PPP\$.<sup>2</sup> Other initiatives typically cover a smaller set of countries. For example, Wang et al. (2013) offers a similar data for a sample of 12 OECD countries obtained from household surveys from 1969 to 2005. The United Nations Economic Commission for Europe (UNECE) Statistical Database compiles from national and international official sources (OECD, Eurostat, CIS), wages and employment information for 52 European countries. Also, the Socio-Economic Database for Latin America (SEDLAC) provides hourly wages in nominal LCUs that can be disaggregated by gender, age, education, economic sector for 24 Latin American countries. Similarly, the U.S. Bureau of Labor Statistics used to produce data for 33 countries as part of the [International Labor Comparisons](#) (ILC) program for assessing U.S. economic and labor market performance relative to other countries.<sup>3</sup>

The International Labour Organization (ILO) compiles labor-specific statistics with a global coverage that is the largest set available. ILO's [ILOSTAT \(formerly LaborStat\)](#) contains annual and semi-annual labor market information for over 230 countries, areas, and territories. Also, its Key Indicators of the Labor Market (KILM) database has country-level data on 18 key indicators (from 1980 onward), which are further disaggregated by industry, gender, and occupation. The GIDD-SWB contributes with these previously available data sources by providing comparable wages in US\$ and PPP\$ for a consistent classification of workers across industries and educational levels. Moreover, it can be anchored to the GTAP data, an information source for over 130 countries used in global CGE modeling.

## 2.1 Constructing the GIDD Sectoral Wage Bill Database

The construction of the GIDD-SWB involves several steps, each of which necessarily has assumptions and caveats: i) the criteria used for the selection of surveys; ii) the information extracted from the household surveys (I2D2); and iii) the estimation of labor skills and wages, including definitions and explanatory notes. The [Global Income Distribution Dynamics](#) website provides more details and additional country-specific statistics derived from this database.

---

<sup>2</sup> The list of surveys used from I2D2 is available in the appendix.

<sup>3</sup> ILC has been discontinued by the Bureau of Labor Statistics but a similar product is now published by [the Conference Board](#). These indicators fall in the following categories: (1) Labor force, employment, and unemployment; (2) Hourly compensation costs (wages and benefits); (3) Productivity and unit labor costs; (4) GDP per capita and per hour; and (5) Consumer prices.

### ***a) Selection of surveys***

We chose 2007 as the base year to split the sectoral wage bill for different skill levels of workers (2007 also matches the baseline year for GTAP 8 database. Unfortunately, not all countries have a usable survey for the base year, and this obliged us to select for each country the closest survey to 2007 that comply with the following criteria:

- Data are available for (i) years of schooling and (ii) industry of employment;
- Survey is not older than 1990;
- When more than one survey is available with a comparable amount of information, we gave priority to the most recent one. This means that if there were surveys for 2008 and 2006, we kept the survey from 2008;
- Similarly, in cases where more than one survey for the same country-year pair existed, we gave priority to nationally representative labor-force surveys.<sup>4</sup>

The approach yielded surveys for 104 countries, covering all relevant regions and income groups as classified by the World Bank (Table 1a). The majority of surveys (85 of them) were recent, about survey years from the period 2005 to 2012; 12 of the surveys correspond to the period 2001 to 2004; and, seven surveys pertain to the years 1992 to 1999.<sup>5</sup> Table 1b and c show that the resulting database has a wide coverage of the underlying population and GDP in the world, spanning approximately 85% of the global population and 80% of the global GDP.<sup>6</sup>

### ***b) The information extracted from the household surveys.***

The I2D2 is the main data source for the construction of the GIDD-SWB database. It is part of a World Bank project to consolidate existing household surveys and harmonize a series of common variables across them. I2D2 collects nationally representative surveys (or samples) on a global scale. In its most current version (version 5, as of June 2014), I2D2 has more than 900 surveys from 1970 to 2013 for more than 150 countries – this vast coverage imposes natural limitations to the set of variables that can be harmonized across time and countries. In the latest version of I2D2, a total of 55 variables are organized into six modules: (i) Database, (ii) Demographic, (iii) Dwelling, (iv) Education, (v) Labor, and (vi) Welfare. Coverage on the variables and quality of data varies by country, region, and time period, and as a result, a significant amount of the work deals with additional survey processing (see Montenegro and Hirn 2009). Table AI.1 in the appendix lists the variables used to generate the GIDD-SWB from the I2D2.

The industry variable identifies the sector of activity of workers. It includes up to 10 sectors, such as agriculture, manufacturing, utilities, construction, and services, consistent

---

<sup>4</sup> The exception is India, where other surveys seem to provide better quality of data.

<sup>5</sup> Additional details and tables are shown in the appendix.

<sup>6</sup> Among the richer countries, Japan is missing.

with the International Standard Industrial Classification of All Economic Activities (ISIC). Table AI.2 in the appendix shows the map between I2D2 and ISIC sectoral classification.

**Table 1 Coverage of the GIDD Sectoral Wage Database**

| Region                                                                                 | HIO               | HIX              | LIC              | LMI               | UMI               | Total              |
|----------------------------------------------------------------------------------------|-------------------|------------------|------------------|-------------------|-------------------|--------------------|
| <b>a) Number of countries with a household survey selected</b>                         |                   |                  |                  |                   |                   |                    |
| East Asia & Pacific                                                                    |                   |                  |                  | 5                 | 2                 | 7                  |
| Europe & Central Asia                                                                  |                   |                  | 1                | 3                 | 7                 | 11                 |
| Latin America & Caribbean                                                              |                   |                  | 1                | 6                 | 11                | 18                 |
| Middle East & North Africa                                                             |                   |                  |                  | 2                 | 2                 | 4                  |
| South Asia                                                                             |                   |                  | 3                | 3                 |                   | 6                  |
| Sub-Saharan Africa                                                                     |                   |                  | 16               | 7                 | 3                 | 26                 |
| High Income                                                                            | 23                | 9                |                  |                   |                   | 32                 |
| <b>Total</b>                                                                           | <b>23</b>         | <b>9</b>         | <b>21</b>        | <b>26</b>         | <b>25</b>         | <b>104</b>         |
| <b>b) Population spanned by the household surveys (in millions)</b>                    |                   |                  |                  |                   |                   |                    |
| East Asia & Pacific                                                                    | -                 | -                | -                | 419               | 1,400             | 1,819              |
| Europe & Central Asia                                                                  | -                 | -                | 7                | 55                | 127               | 189                |
| Latin America & Caribbean                                                              | -                 | -                | 9                | 36                | 448               | 493                |
| Middle East & North Africa                                                             | -                 | -                | -                | 94                | 14                | 108                |
| South Asia                                                                             | -                 | -                | 199              | 1,340             | -                 | 1,539              |
| Sub-Saharan Africa                                                                     | -                 | -                | 374              | 78                | 69                | 521                |
| High Income                                                                            | 779               | 162              | -                | -                 | -                 | 941                |
| <b>Total</b>                                                                           | <b>779</b>        | <b>162</b>       | <b>590</b>       | <b>2,021</b>      | <b>2,058</b>      | <b>5,609</b>       |
| <b>c) Percent share of GDP covered</b><br>(figure in parenthesis is the missing share) |                   |                  |                  |                   |                   |                    |
| East Asia & Pacific                                                                    | -                 | -                | -                | 3.0               | 12.5 (0.6)        | 15.5 (0.7)         |
| Europe & Central Asia                                                                  | -                 | -                | -                | 0.5 (0.2)         | 2.4 (0.7)         | 3.0 (0.9)          |
| High Income                                                                            | 40.6 (9.5)        | 4.3 (3.6)        | -                | -                 | -                 | 44.9 (13.0)        |
| Latin America & Caribbean                                                              | -                 | -                | -                | 0.2 (0.1)         | 6.4 (0.9)         | 6.7 (1.0)          |
| Middle East & North Africa                                                             | -                 | -                | -                | 0.9 (0.3)         | 0.2 (2.3)         | 1.1 (2.6)          |
| South Asia                                                                             | -                 | -                | 0.4              | 7.4               | -                 | 7.8                |
| Sub-Saharan Africa                                                                     | -                 | -                | 0.5 (0.1)        | 0.2 (1.1)         | 0.9 (0.1)         | 1.6 (1.2)          |
| <b>Total</b>                                                                           | <b>40.6 (9.5)</b> | <b>4.3 (3.6)</b> | <b>0.9 (0.2)</b> | <b>12.3 (1.7)</b> | <b>22.4 (4.6)</b> | <b>80.5 (19.5)</b> |

Source: GIDD Sectoral Wage Bill Database

Note: HIO: High-Income Countries (OECD); HIX: High-Income Countries (non-OECD); LIC: Low-Income Countries; LMI: Lower Middle-Income; UMI: Upper Middle-Income.

### *c) Labor skills and wages*

The current version of the GIDD-SWB defines two types of workers based on educational attainment, namely workers with 0-9 (unskilled), and more than ten years of schooling (skilled). Nonetheless, the database can be expanded to add cohorts using different schooling thresholds (e.g. 0-4, 5-9, 10-12 and above 12 years of schooling). Although refinement in the

number skill categories may improve the analysis of human capital formation in some ways, it could, however, limit in other ways. For example, more categories of education/skill may not capture the improvement of education in less educated countries, relative to the rest of the world.

Following standard definition, the classification of workers pertains to “employed” individuals between 15 and 64 years old (working-age population), excluding unpaid employees since the latter would not be accounted in the total wage bill (their wages are equal to zero). Whenever a selected survey is not from 2007, a scaling adjustment is made equally to all sampling weights to reach the working-age population aggregate for 2007 in the UN database while keeping the structure of the household survey.<sup>7</sup> For the large majority of surveys, population size and number of workers are consistent with the information publicly available on sources such as the World Development Indicators (WDI). The resultant data covers approximately 2.1 billion of workers, which are widely distributed across different industry and income groups of countries. A large majority of workers in agriculture belongs to the upper-middle, lower-middle, and low-income level countries. Unskilled workers in agriculture still represent a large share of workers in low-income countries.

For most countries, the *wage* variable in the I2D2 household surveys is available in nominal local currency units (LCUs) and different for each worker. For a few countries without wage information, we use income or expenditure per worker as a proxy for the relevant *wage*, which was defined as the total income (or expenditure) divided by the total number of workers in the household. Because important discrepancies can occur between wages and income or expenditure per worker, the option to not use proxy measures is provided. All told, the GIDD-SWB employs wage information for 89 countries, per worker income for four countries (Liberia, Mauritania, Tukey, Ukraine), and per worker consumption in eight countries (Angola, Georgia, Indonesia, India, Niger, Russia, Senegal, and Swaziland). Three countries (Guinea, Lebanon, and Lesotho) only have the volume or number of workers available.

LCU values of wages are converted to US dollars and international dollars using US\$/LCUs and PPP\$/LCUs exchange rates, respectively. If a selected survey is not from 2007, national consumer price index ratios (between each survey year and 2007) are used to convert all money values into real 2007 LCUs. Similarly, real GDP per capita ratios are applied to account for real wage increases between periods.<sup>8</sup>

---

<sup>7</sup> We used the UN (2012) as reference for population projections.

<sup>8</sup> In order to calculate average wages by country, industry, skill level, and type of worker, we simply divide the wage bill by the number of workers. It is important to emphasize that scattered missing information is common across household surveys and I2D2 is not an exception. Some individuals reporting wages lack information on their industry of employment, and vice-versa. For the purpose of constructing the GIDD-SWB, in those cases where industry is missing, no imputation has been done and average wages are reported with a missing code for industry. In those cases where there is missing information on wages, the procedure used implies that average sectoral wages are attributed to individuals with missing information on wages. In other words, the average wage takes into account only the non-missing information.

Finally, to avoid the possibility of having outliers driving the results we exclude<sup>9</sup> 0.5% of the observations in the upper and 0.5% of the observation in the lower wage distribution. However, when analyzing inequality this procedure might entail two offsetting effects: although income may be strongly concentrated in the top 1% of the distribution, it might also be underreported.<sup>10</sup>

## 2.2 Results and stylized facts

The database provides information about labor and wages for 104 countries and splits the value added of labor in key economic sectors into employment levels and wages according to skilled and unskilled workers. In so doing, it derives the indicative magnitude of the skill premium of wage with a consistent definition across countries. Appendix II summarizes the average wage, number of workers by skill, industry, and broad level of income for 2007. Wages are also provided in US\$ and PPP\$. The data for specific countries are available in the database.

### a) Wage Skill premia

Table 2 reports the wage skill premia of wages as the ratios of average wages between skilled workers and unskilled workers, by sector, and income group of countries. Because countries belonging to the particular income groups may have different relative weights regarding PPP and US\$, the skill premia may differ between PPP and US\$. Such difference does not exist for an individual country. Consistent with the findings of Acemoglu (2003), Acemoglu and Autor (2012) and Berman (2013), a larger share of skilled workers is positively correlated with the level of skill premia. Hence, skill premia of wages are relatively larger across several sectors in high-income countries.

**Table 2 Skill premia of wages by industry and income level  
(Wage of unskilled worker = 1.0)**

| Industry         | Wages, PPP\$(2007) |      |      |      |      | Wages, US\$(2007) |      |      |      |      |
|------------------|--------------------|------|------|------|------|-------------------|------|------|------|------|
|                  | HIO                | HIX  | LIC  | LMI  | UMI  | HIO               | HIX  | LIC  | LMI  | UMI  |
| Agriculture      | 2.26               | 1.35 | 1.44 | 1.56 | 1.55 | 2.08              | 1.36 | 1.11 | 1.50 | 1.63 |
| Mining           | 2.01               | 1.37 | 1.22 | 2.99 | 1.65 | 1.99              | 1.37 | 1.18 | 2.58 | 1.65 |
| Manufacturing    | 2.55               | 1.39 | 1.84 | 2.31 | 1.74 | 2.04              | 1.43 | 1.76 | 2.03 | 1.83 |
| Public utilities | 2.52               | 1.72 | 2.30 | 1.49 | 1.47 | 2.52              | 3.78 | 2.10 | 1.55 | 1.71 |
| Construction     | 2.08               | 1.30 | 1.65 | 1.59 | 2.19 | 1.77              | 1.13 | 1.64 | 1.47 | 2.30 |
| Commerce         | 1.99               | 1.52 | 1.38 | 1.50 | 1.94 | 1.60              | 1.64 | 1.13 | 1.35 | 1.88 |

<sup>9</sup> Actually, we attributed missing values to the wages of those individuals.

<sup>10</sup> In the appendix we provide some descriptive statistics on the excluded outliers.

|                        |      |      |      |      |      |      |      |      |      |      |
|------------------------|------|------|------|------|------|------|------|------|------|------|
| Transports and commun. | 1.89 | 1.20 | 2.10 | 1.59 | 1.54 | 1.54 | 1.20 | 2.00 | 1.40 | 1.53 |
| Financial and business | 2.78 | 3.32 | 3.33 | 1.76 | 2.08 | 2.30 | 4.82 | 2.56 | 1.52 | 2.05 |
| Community services     | 2.45 | 1.67 | 1.99 | 1.82 | 2.30 | 1.96 | 2.06 | 1.45 | 1.46 | 2.27 |
| Others                 | 1.73 | 3.20 | 2.20 | 1.61 | 2.20 | 1.73 | 4.69 | 2.11 | 1.66 | 2.57 |

Source: GIDD Sectoral Wage Bill Database

Note: HIO: High-Income Countries (OECD); HIX: High-Income Countries (non-OECD); LIC: Low-Income Countries; LMI: Lower Middle-Income; UMI: Upper Middle-Income.

## b) Distribution of workers and wages by skill.

Table 3 and 4 summarize the wage share and distribution of workers by skill, industry, and broad income level.

**Table 3 Wage share (in USD) by type of worker, industry, and income level**

| Industry               | Wage share of skilled workers |     |     |     |     | Wage share of unskilled workers |     |     |     |     |
|------------------------|-------------------------------|-----|-----|-----|-----|---------------------------------|-----|-----|-----|-----|
|                        | HIO                           | HIX | LIC | LMI | UMI | HIO                             | HIX | LIC | LMI | UMI |
| Agriculture            |                               |     |     |     |     |                                 |     |     |     |     |
| Mining                 |                               |     |     |     |     |                                 |     |     |     |     |
| Manufacturing          |                               |     |     |     |     |                                 |     |     |     |     |
| Public utilities       |                               |     |     |     |     |                                 |     |     |     |     |
| Construction           |                               |     |     |     |     |                                 |     |     |     |     |
| Commerce               |                               |     |     |     |     |                                 |     |     |     |     |
| Transports and commun. |                               |     |     |     |     |                                 |     |     |     |     |
| Financial and business |                               |     |     |     |     |                                 |     |     |     |     |
| Community services     |                               |     |     |     |     |                                 |     |     |     |     |
| Others                 |                               |     |     |     |     |                                 |     |     |     |     |

Source: GIDD Sectoral Wage Bill Database

Note: HIO: High-Income Countries (OECD); HIX: High-Income Countries (non-OECD); LIC: Low-Income Countries; LMI: Lower Middle-Income; UMI: Upper Middle-Income.

**Table 4 Distribution of the number of workers by skill, industry, and income level**

| Industry               | Share of skilled workers |     |     |     |     | Share of unskilled workers |     |     |     |     |
|------------------------|--------------------------|-----|-----|-----|-----|----------------------------|-----|-----|-----|-----|
|                        | HIO                      | HIX | LIC | LMI | UMI | HIO                        | HIX | LIC | LMI | UMI |
| Agriculture            |                          |     |     |     |     |                            |     |     |     |     |
| Mining                 |                          |     |     |     |     |                            |     |     |     |     |
| Manufacturing          |                          |     |     |     |     |                            |     |     |     |     |
| Public utilities       |                          |     |     |     |     |                            |     |     |     |     |
| Construction           |                          |     |     |     |     |                            |     |     |     |     |
| Commerce               |                          |     |     |     |     |                            |     |     |     |     |
| Transports and commun. |                          |     |     |     |     |                            |     |     |     |     |
| Financial and business |                          |     |     |     |     |                            |     |     |     |     |
| Community services     |                          |     |     |     |     |                            |     |     |     |     |
| Others                 |                          |     |     |     |     |                            |     |     |     |     |

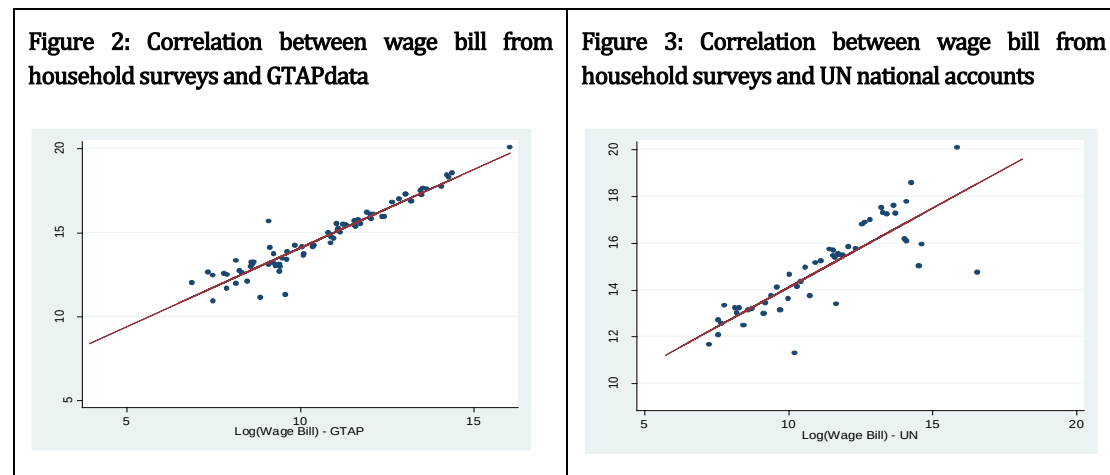
Source: GIDD Sectoral Wage Bill Database

Note: HIO: High-Income Countries (OECD); HIX: High-Income Countries (non-OECD); LIC: Low-Income Countries; LMI: Lower Middle-Income; UMI: Upper Middle-Income.

c) Consistency with national accounts and other data sources.

a) National account

A dataset on wage bill by sector of activity with global coverage, generated from household surveys, opens a lot of possibilities of analysis linking macro and micro perspectives. However, is the wage bill from household surveys consistent with the national accounts? In other words, do the wages from household data sum up to the aggregate and sectoral value added of labor found in national accounts or SAMs? Recent efforts checking the consistency between household surveys and national accounts (e.g. Deaton 2010 and Bourguignon et. al. 2008) found significant differences between these values, which can affect poverty measurement (see the case of India in Deaton and Kozel 2005). Indeed, we expect the levels to be different for various reasons, such as recall period (seasonality, inflation), or mixed incomes (self-employed). Another potential explanation is that high-income individuals are underrepresented in the surveys or their incomes are under reported.



Source: authors' calculations.

Figure 2 compares the total value of wage bill (in thousands US\$ of 2007) from GIDD-SWB and the GTAP database, which provides standardized SAMs or input-output tables for many countries. Overall, we observe a positive correlation between both sources of information. Figure 3 shows a similar information comparing the total value of wage bill from GIDD-SWB with the UN national accounts. The two databases are highly positive correlated. Although we found discrepancies regarding values (in level) for most of the countries, the ranking follows similar or consistent patterns. This observation has implications to how information about labor income should be exchanged between the macro and micro level of the GIDD framework (see below).

b) Wages in US\$ and PPP\$.

Table 5 shows the ratio of the average wage of each industry by country's income level relative to the low-income countries (which are set to 1.0s). As expected, the ratios in PPP\$ are much lower than those in US\$. Even so, the difference in average wage between low-income countries and high-income OECD is still quite dramatic in PPP\$. For example, the wages of unskilled workers in PPP\$ for manufacturing in HIO countries can be ten times larger than those in low-income countries; similarly, they are about 14 times larger for skilled workers.

### c) Wages and GDP per capita

We observe a positive and strong correlation between average wage and GDP per capita of approximately 67% (Figure 4). The same is observed when we split the economy broadly into agriculture and non-agriculture activities (Figure 5). The figures show a wider dispersion of the average wage among countries with a lower GDP per capita. Noting that labor income is likely the main source of earnings for most of the population, these figures seem to confirm a large labor income inequality "between countries." Furthermore, labor factor remuneration seems to vary significantly among countries (i.e. many countries with very different GDP per capita have relatively similar average wage).

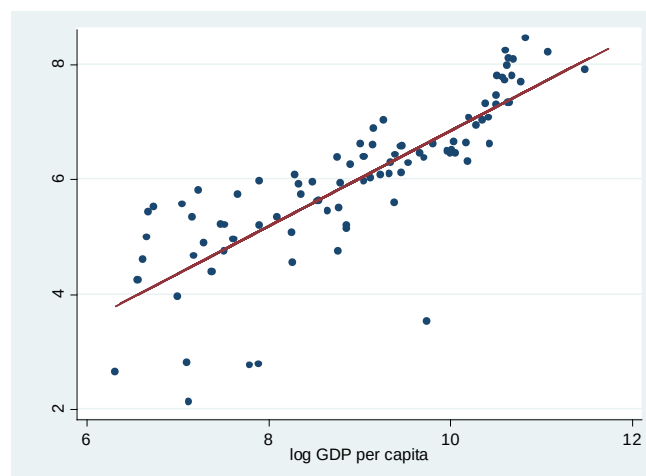
**Table 5 Ratio of average wage between groups of countries**

| Skill level | Industry                 | Wages, PPP\$(2007) |     |     |     |     | Wages, US\$(2007) |      |     |     |     |
|-------------|--------------------------|--------------------|-----|-----|-----|-----|-------------------|------|-----|-----|-----|
|             |                          | HIO                | HIX | LIC | LMI | UMI | HIO               | HIX  | LIC | LMI | UMI |
| Unskilled   | Agriculture              | 7.5                | 4.0 | 1.0 | 1.1 | 0.9 | 13.2              | 3.9  | 1.0 | 0.7 | 0.9 |
|             | Mining                   | 15.3               | 3.1 | 1.0 | 0.8 | 2.2 | 19.0              | 2.8  | 1.0 | 0.6 | 2.1 |
|             | Manufacturing            | 10.2               | 3.9 | 1.0 | 1.1 | 1.7 | 27.2              | 5.3  | 1.0 | 1.2 | 2.6 |
|             | Public utilities         | 18.4               | 5.3 | 1.0 | 1.9 | 2.3 | 33.7              | 9.8  | 1.0 | 1.7 | 3.3 |
|             | Construction             | 8.9                | 3.3 | 1.0 | 1.0 | 1.2 | 25.2              | 10.6 | 1.0 | 1.1 | 1.9 |
|             | Commerce                 | 6.7                | 2.3 | 1.0 | 1.0 | 1.4 | 11.7              | 4.0  | 1.0 | 0.8 | 1.4 |
|             | Transports and communic. | 12.0               | 4.9 | 1.0 | 1.4 | 2.2 | 30.4              | 11.5 | 1.0 | 1.4 | 3.2 |
|             | Financial and business   | 13.3               | 3.3 | 1.0 | 1.8 | 3.0 | 30.0              | 6.2  | 1.0 | 1.8 | 4.1 |
|             | Community services       | 6.8                | 2.4 | 1.0 | 1.2 | 1.2 | 17.1              | 5.7  | 1.0 | 1.3 | 1.8 |
|             | Others                   | 7.5                | 1.9 | 1.0 | 1.3 | 1.1 | 25.1              | 5.2  | 1.0 | 1.4 | 2.0 |
| Skilled     | Agriculture              | 11.8               | 3.8 | 1.0 | 1.2 | 1.0 | 24.9              | 4.7  | 1.0 | 1.0 | 1.3 |
|             | Mining                   | 25.2               | 3.5 | 1.0 | 2.1 | 2.9 | 32.0              | 3.2  | 1.0 | 1.3 | 2.9 |
|             | Manufacturing            | 14.1               | 2.9 | 1.0 | 1.4 | 1.6 | 31.6              | 4.3  | 1.0 | 1.4 | 2.7 |
|             | Public utilities         | 20.1               | 4.0 | 1.0 | 1.2 | 1.5 | 40.4              | 17.6 | 1.0 | 1.3 | 2.7 |
|             | Construction             | 11.2               | 2.6 | 1.0 | 1.0 | 1.6 | 27.2              | 7.3  | 1.0 | 1.0 | 2.7 |
|             | Commerce                 | 9.7                | 2.5 | 1.0 | 1.1 | 2.0 | 16.5              | 5.9  | 1.0 | 0.9 | 2.4 |
|             | Transports and communic. | 10.8               | 2.8 | 1.0 | 1.1 | 1.6 | 23.3              | 6.9  | 1.0 | 1.0 | 2.4 |
|             | Financial and business   | 11.1               | 3.3 | 1.0 | 1.0 | 1.8 | 27.0              | 11.7 | 1.0 | 1.1 | 3.3 |
|             | Community services       | 8.3                | 2.0 | 1.0 | 1.1 | 1.4 | 23.1              | 8.0  | 1.0 | 1.3 | 2.7 |
|             | Others                   | 5.9                | 2.7 | 1.0 | 0.9 | 1.1 | 20.6              | 11.5 | 1.0 | 1.1 | 2.4 |

Source: Authors' calculations from GIDD Sectoral Wage Bill Database

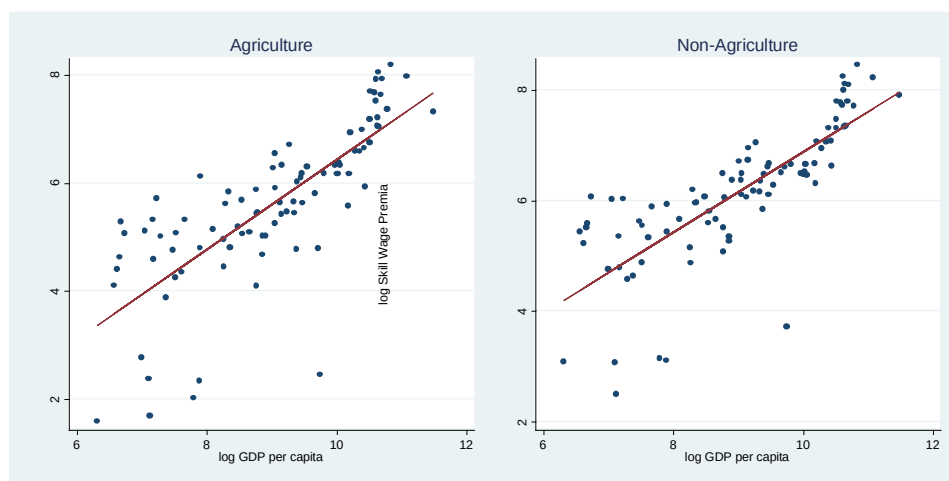
Note: HIO: High-Income Countries (OECD); HIX: High-Income Countries (non-OECD); LIC: Low-Income Countries; LMI: Lower Middle-Income; UMI: Upper Middle-Income.

**Figure 4 Correlation between wages and GDP per capita**



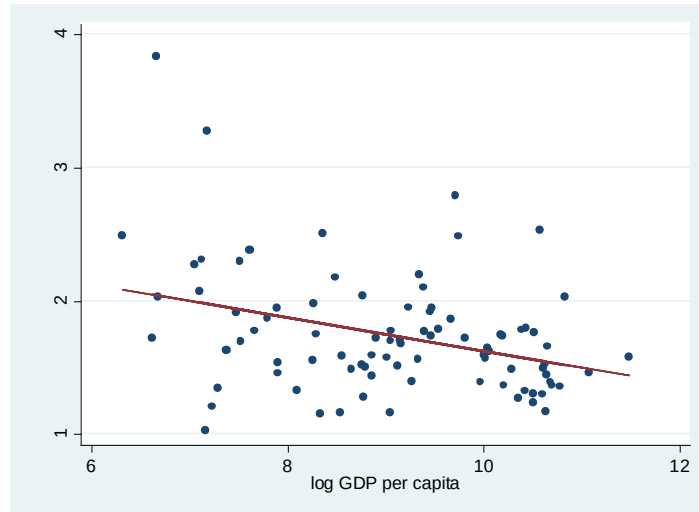
Source: Authors' calculations.

**Figure 5 Correlation between sectoral wages and GDP per capita**

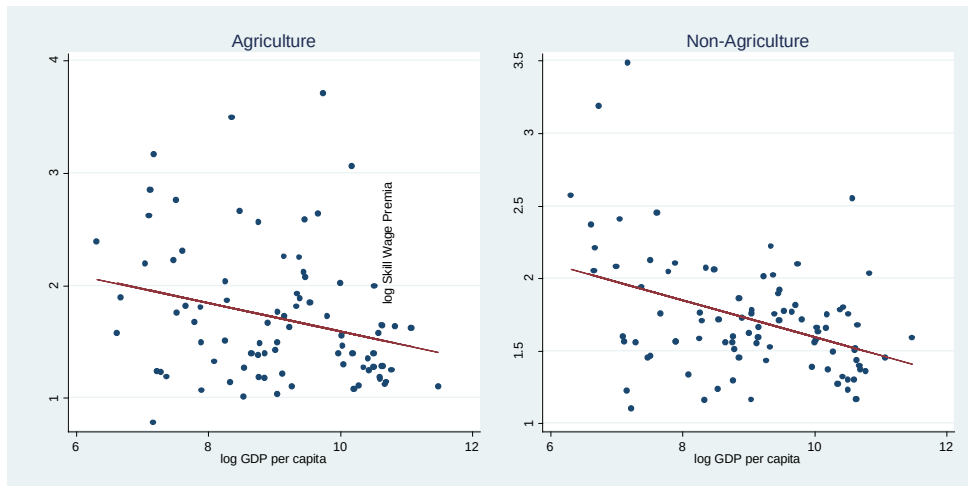


Source: Authors' calculations.

**Figure 6 Correlation between wage skill premia and GDP per capita**



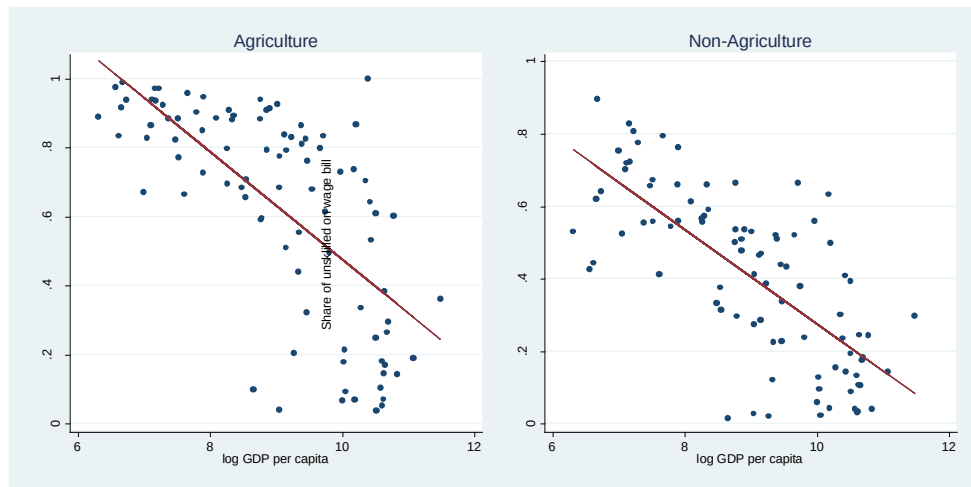
**Figure 7 Correlation between wage skill premia and GDP per capita**



***Skill wage premia and GDP per capita.*** On the other hand, we observe a negative correlation between skill wage premia and GDP per capita. Even so, the correlation is relatively weak (about 12%), but the coefficient is negative and statistically significant. Also, Figure 6 and 7 show a large variance for skill premia at the same level of GDP per capita. Although these numbers suggest a slight reduction in skill premia when countries get richer, Autor (2014) suggests that the skill wage premia has increased in a large number of high-income economies in the last decades, contributing to increasing earnings inequality in these countries.

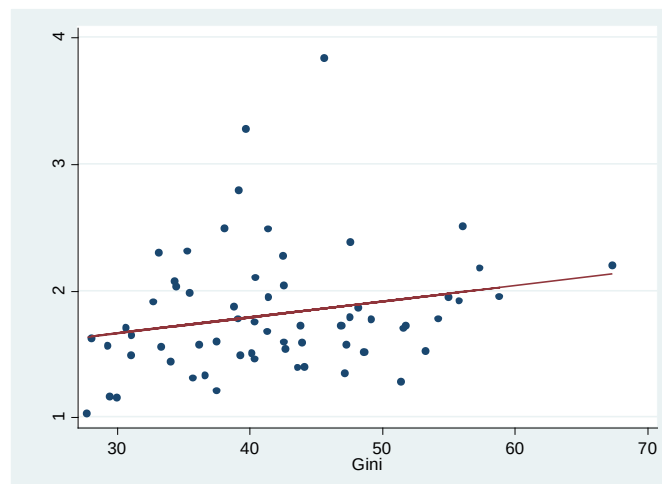
***Wage skill premia and other factors.*** Figure also shows a negative correlation between the share of unskilled workers in the wage bill and GDP per capita. This finding is relatively similar for agriculture and non-agriculture activities.

**Figure 8 Correlation between share of unskilled workers on wage bill and GDP per capita**

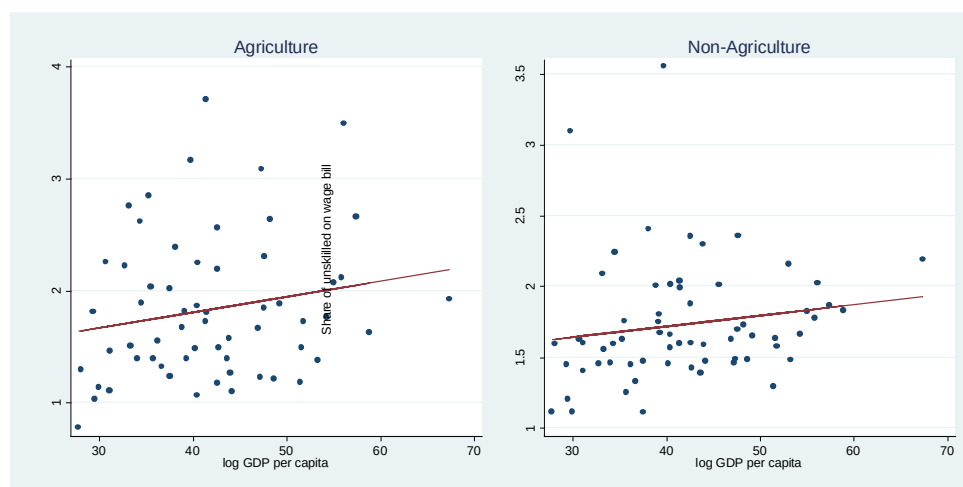


Finally, Figure 9 shows a weak positive correlation between skill wage premia and the Gini coefficient. There is a large dispersion in the Gini for the same level of average wage. The same is observed for the sectoral wage of agriculture and non-agriculture activities against the Gini (Figure 10).

**Figure 9 Correlation between skill wage premia and Gini**



**Figure 10: Correlation between skill wage premia and Gini by sector (wrong graph – see axis labels)**



### 3 Microsimulation with GIDD 2.0

With GIDD-SWB database, an important feature of GIDD 2.0 is the accounting of relative changes in the supply of skilled and unskilled workers over time using education attainment to differentiate workers by level of skill. Such human capital formation plays a key role in economic growth, poverty and income distribution. In particular, the set of Linkage Aggregate Variables (LAVs) connecting the macro-micro levels has several dimensions. It combines geographical aggregation (global, regional groups, or individual country), time dimensions, demographic structure (age groups, gender, and level of education), labor characteristics (participation status, type of qualification by education), factor income (wage, non-wage), sector of employment, welfare aggregate (income/consumption per capita), and price index (as defined by consumption bundle).

Incorporating demographic trends and changes in the skill composition of labor supply into the macro-micro simulation involves the following broad steps:

1. Defining scenarios for demographic change and skill accumulation of labor supply
2. Setting the base-year or starting household sampling weights from the household survey.
3. Estimating a new set of household sampling weights to replicate the changes in the demographic structure over time as defined by the baseline and counterfactual scenarios.
4. Accounting for labor migration from a shrinking to an expanding economic sector.

5. Changes in sectoral wage premia are modeled.
6. In step four, distribution-neutral growth is applied.
7. Lastly, changes in relative consumption prices reflect changes in real incomes.

Appendix III offers greater details about the methodology, including the main symbols and equations. Here, we summarize the key steps – the population and education projections, the mathematical problem of solving for the new household sampling weights.

### 3.1 Population and education projections

To take advantage of GIDD’s new features, a forward-looking exercise requires population projections that differentiate workers by level of skill, usually using education attainment as a proxy. In this regard, GIDD makes use of two sources of population projection, the UN-WPP (2012) and the IASA (2012). These projections use the cohort component method, which generates population trends by age groups and other demographic attributes (e.g. gender). The uncertainty regarding these projections pertains to assumptions about fertility, mortality, and migration rates (Cohen, 2001; Lutz, 2001).<sup>11</sup>

The UN-WPP (2012) projection incorporates the fact that every year there is a new cohort of younger workers that replace an older cohort in the working age population, a “pipeline effect. If the share of skilled workers among the younger cohorts is different from that of the older cohorts, the pipeline effect will change the skill composition of workers over time. The additional benefit of using this option is the consistency between the projection and the initial demographic distribution observed in the household surveys used in GIDD. Because of the natural progression of children through the school system as they age, the pipeline effect on the skill composition of labor will exist even if school enrollment rates do not change. To capture the impact of policy improvements in the education system over time, additional assumptions are necessary.

The IASA (2012) case has the advantage of incorporating educational attainment in its projection. It also addresses the fact that educational attainment may affect fertility over time. Even so, the initial share of skilled and unskilled workers may not match with the household surveys, which can be a drawback.

#### a) Re-weighting the household surveys

The next step in the microsimulation is to establish a common set of pathways for the demographic structure in both the macro and micro models. Once the alternative demographic pathways are exogenously determined, the objective is to obtain a set of

---

<sup>11</sup> For a summary on the literature related to accuracy of population forecast, see Cruz (2015).

household sampling weights that replicate the counterfactual demographic structure using the minimum cross-entropy method proposed by Wittenberg (2010). This method allows the recreation of homogeneous weights within household members. In most household surveys, the unit of analysis and sampling is the household; hence, sampling weights attached to every individual within the household are equal.

The basic mathematical problem is to derive the new household weight ( $\tilde{\theta}_{ht_1}^*$ ) or its probability distribution ( $\tilde{\vartheta}_{ht_1}^*$ ) from the initial weight ( $\theta_{ht_0}^*$ ) or distribution ( $\vartheta_{ht_0}^*$ ) and by imposing several constraints such as education, gender, or age-group population growth rates. Formally, the new probability distribution can be obtained by solving the optimization problem in equation (1):

$$\min \Gamma(\vartheta, \tilde{\vartheta}) = \sum_h \tilde{\vartheta}_{ht_1}^* \ln\left(\frac{\tilde{\vartheta}_{ht_1}^*}{\vartheta_{ht_0}^*}\right) \quad (1)$$

subject to

$$\begin{aligned} E(x_{rt}) &= \sum_h \vartheta_{ht_0}^* k_{x_{rt}h} \quad r=1, \dots, R \\ \sum_h \vartheta_{ht_0}^* &= 1 \end{aligned}$$

The term  $k_{x_{rt}h}$  is the mean of the  $x_r$  characteristic within the household  $h$  ( $k_{x_{rt}h} = \frac{\sum_i x_{iht}}{hsize_h}$ ). In other words, this optimization procedure picks the best distribution  $\tilde{\vartheta}_{ht_1}^*$  that meets the moment constraints and the normalization restriction while, at the same time, requiring the least additional information and the least distance from the initial distribution ( $\vartheta_{ht_0}^*$ ). The solution to this problem is given by equation (2):

$$\begin{aligned} \hat{\vartheta}_{ht_1} &= \vartheta_{ht_0}^* \frac{\exp(\chi \hat{\lambda})}{\Omega(\hat{\lambda})} \rightarrow \sum_h \hat{\vartheta}_{ht_1} = \sum_h \vartheta_{ht_0}^* q_h = \tilde{H}_{t_1} \\ q_h &= \frac{\exp(X \hat{\lambda})}{\Omega(\hat{\lambda})} \end{aligned} \quad (2)$$

in which  $\hat{\lambda}$  is the Lagrange multiplier for the constraint  $E(X)$ , and  $\Omega$  is the normalizing factor to scale the new population to the target size.

b) Other steps

**The reallocation of labor.** To account for the prospects of labor movements across economic sectors and over time as defined by the scenarios in the macro or CGE model, the GIDD implements a probabilistic way to select the migrants from the shrinking sector to move

into the expanding sector. The probability of observing an individual  $i$  working in the expanding sector ( $p=p_1$ ) can be modeled as a probit equation (3):

$$Pr(p = p_1) = P(X_i, Z_i) \quad (3)$$

where  $X_i$  and  $Z_i$  are vectors of individual and household characteristics, respectively, for individual  $i$ . Workers in their shrinking sectors,  $p_0$ , can be assigned a score based on their set of characteristics  $X$  and  $Z$ , and sorted accordingly. Workers in the shrinking sector,  $p_0$ , with the higher probability of being in the expanding sector,  $p_1$ , can migrate up to a point where the predicted share of labor is satisfied. That predicted share is a result of the macro model.

**Changes in wages and skill premia.** The next step adjusts factor returns by skill and sector by the results of the CGE model. The GIDD imposes an entirely new vector of earnings on each worker, conditional on that worker being in sector  $p$  and qualification level  $q$ . To ensure consistency between the macro and micro outcomes, the GIDD adjusts the ratios between wage premia rather than wages themselves. This procedure implies that even if initial and final wages differ between the macro and micro models, the change in wage ratios will be consistent across the two models. Such approach diminishes the possibility of a reversal in the wage gap between skill levels and ensures that the distributional changes are consistently mapped from the macro to the microdata.

**Accounting for economic growth.** In this step, we adjust the microdata such that the weighted average percentage change in the per capita income/consumption across all households matches the change in real consumption per capita in the CGE model. We implement this constraint in a distribution-neutral way. That is, we adjust all household income in the counterfactual by a scalar  $\lambda$ :

$$Y' \frac{\hat{Y}}{Y} = \hat{Y}' = \lambda \left\{ \sum_h \omega_h \hat{Y}'_h \right\} \quad (4)$$

The ratio  $\frac{\hat{Y}}{Y}$  is the growth factor of income between the base-year real per capita income,  $Y$ , and the predicted real per capita income,  $\hat{Y}$ , from the macro or CGE model. That growth factor is applied to the old *weighted* average of household income  $Y'$  get its new value  $\hat{Y}'$ . The new income for each household  $\hat{Y}'_h$  is then multiplied by the scalar  $\lambda$  using the new household weight  $\omega_h$  in the policy simulation, such that  $\sum_h \omega_h = 1$ .

This step in the microsimulation assures that national income grows accordingly to the results of the CGE model. Despite the fact that the GIDD ignores other forms of income, such as capital income (primarily because of missing data), this transformation guarantees consistency between the weighted average income in the household and CGE model assessment. For poor households, which is the focus of the work, the assumption should be reasonably accurate since poor households have little capital income. There is more of a margin of error for wealthier households. Even so, Bussolo, de Hoyos, Medvedev (2010) noted the tendency for skilled wage and the return to capital to be correlated.

**Accounting for food and non-food prices.** Given the importance of food consumption in the household budget for many developing countries, the GIDD accounts for the effects of food and non-food prices in household expenditures. As data improves in the household survey, more categories of consumption could be added.

The household-specific consumer price index  $P_h$  is a weighted average of the food price index  $P_f$  and the non-food price index  $P_{nf}$ , with  $\alpha_h$  as the budget share for food (that is,  $P_h = \alpha_h P_f + (1 - \alpha_h) P_{nf}$ ). The parameter  $\alpha_h$  is estimated from the following regression:

$$\alpha_h = \beta_0 + \beta_1 \ln(\gamma'_h) + e_h \quad (4)$$

where  $e_h$  is a vector of household-specific errors that are assumed to be distributed with  $E(e_h) = 0$  and  $V(e_h) = \sigma^2$ . Assuming that estimated parameters  $\hat{\beta}_0$  and  $\hat{\beta}_1$  remain constant, the new budget share spent on food  $\alpha'_h$  for household  $h$  at the counterfactual per capita income,  $\lambda \hat{\gamma}'_h$ , can be obtained from:

$$\alpha'_h = \hat{\beta}_0 + \hat{\beta}_1 \ln(\lambda \hat{\gamma}'_h) + \hat{e}_h \quad (1)$$

The changes in real per capita incomes brought about by a change in prices can be approximated by the following linear expression:

$$\hat{\gamma}_h^r = \frac{\lambda \hat{\gamma}'_h}{\alpha'_h P'_f + (1 - \alpha'_h) P'_{nf}} \quad (2)$$

where  $\hat{\gamma}_h^r$  is the real per capita income, which is also the counterfactual welfare aggregate for the analysis of poverty and shared prosperity.

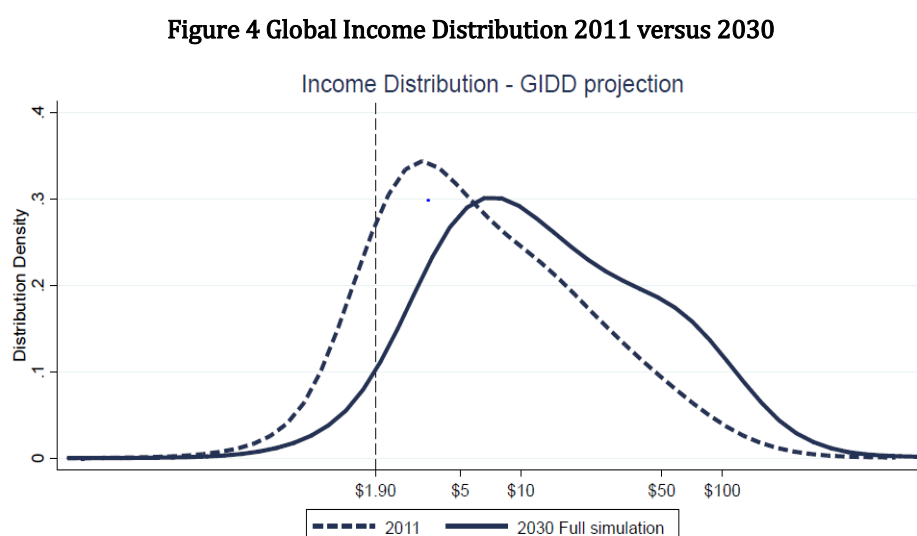
## 4 Illustrations

Among the several applications mentioned in the introduction, we summarized two cases as illustrations: i) the global impact of the coming wave of educated workers from developing countries in Ahmed et al. (2017), and ii) the future threat of the antimicrobial resistance in World Bank (2016).

### 4.1 Education wave

The first case considers that, in the decades ahead, younger and better-educated cohorts will enter the workforce while older, less educated, ones leave. This rising level of schooling, especially in developing countries, will alter the skill composition of labor supply as well as income distribution in the world. To what extent? Using the framework just described with a global general equilibrium framework, the findings under alternative assumptions suggest that global income inequality will likely decrease by 2030. This outcome is mostly driven by shrinking income gaps among countries. However, unlike the increase in within-country inequality from the recent globalization of the world's labor market, the education wave would minimize the further increase of within-country inequality.

Figure 4 shows the effect on global income distribution in 2030 compared to 2011. The 2030 curve is the income distribution associated with the full impact of the education wave in the baseline case. It is not distribution-neutral relative to 2011 curve because it has the full impact of several factors that alter the shape of the curve. First, it incorporates the demographic effect that changes the characteristics of households in 2030, requiring the reweighting of the household sample. That is, individuals that become more representative in the projected population for 2030 (according to their age cohort, gender, and school attainment) have their weight increased. The second channel of changes comes from variations in relative wages and sectoral employment (e.g. due to workers moving from agriculture to non-agricultural sectors). Finally, because countries and regions are growing at different rates, the effect of economic growth on household incomes is not necessarily neutral with respect to global income distribution - the between-country effect. All told, the 2030 curve seems slightly wider, and the slope of the kernel density function in 2030 is not as steep as in 2011 when it crosses the poverty line (1.9 daily international dollars under PPP), meaning fewer people remains in extreme poverty. But it also suggests that poverty reduction might become less sensitive to economic growth around the threshold of the poverty line. Yet, the education wave will benefit the target of reducing the percentage of people living in extreme poverty to less than 3% by 2030.



Source: Ahmed et al. (2017).

The microsimulation results confirm that the world will likely become more equal by 2030 in the baseline scenario (see Table 6, first three columns). The (individual-based) Gini index falls from 64.2 in 2011 to 62.9 in 2030, while the Theil-L index reduced from 83 to 75.5. Compared to recent patterns, these results suggest a continuation of reductions in global inequality. In the first wave, global inequality decreased by 2.4 percentage points in a 20-year interval from 1988 to 2008 (Lakner and Milanovic 2016), our baseline scenario shows a comparable, but slightly smaller, reduction in the second wave of 2.1 percentage point, with the global Gini index falling the period 2011 up to 2030. These results are compatible with expected slowdown of large emerging markets, particularly China (World Bank 2015b). The Theil-L index also suggests a compatible, but smaller, reduction of 7.5 points by 2030, instead of a reduction of 11.5 points observed in the first wave. Compared to Lakner and Milanovic (2016), our initial measure of Gini and Theil-L indices are smaller, which may be explained by differences in benchmark year, data sources, country coverage, and the slight improvement in equality when the second wave is measured by 2011.<sup>12</sup>

Using the Theil index, it is possible to decompose inequality between and within countries. Given the convergence in per capita income, the results confirm that the reduction in global inequality is mostly driven by the component between countries. Between the initial year (2011) and 2030 in the baseline scenario, the contribution of inequality between countries decreases significantly by about 20% (from 48 to 38.6) while the within-country contribution increases by about 5% (from 35 to 36.9). These results are also comparable to the results in Lakner and Milanovic (2016), where they observe a reduction in the contribution of the between-country component to global inequality from 80.8 in 1988 to 72.9 in 2008.<sup>13</sup>

**Table 6 Effect of education wave on global inequality (baseline scenario)**

| Inequality measures             | 2011 | 2030 – education wave |          | 2030 -<br>No<br>Education<br>Wave |
|---------------------------------|------|-----------------------|----------|-----------------------------------|
|                                 |      | Demographic           | Baseline |                                   |
| Gini index (%)                  | 64.2 | 64.2                  | 62.9     | 63.4                              |
| Theil-L (%)                     | 83.0 | 83.8                  | 75.5     | 76.8                              |
| Between - Theil, by region (%)  | 42.8 | 41.8                  | 31.2     | 32.2                              |
| Within - Theil, by region (%)   | 40.1 | 42.0                  | 44.2     | 44.5                              |
| Between - Theil, by country (%) | 48.0 | 47.1                  | 38.6     | 39.3                              |
| Within - Theil, by country (%)  | 35.0 | 36.8                  | 36.9     | 37.4                              |

<sup>12</sup> The latest year available in Lakner and Milanovic (2016) is 2008 for which the Gini is 67.0, based on the 2011 PPP exchange rates, and the Theil-L index is 88.6 (table A3). These values are smaller than those based on the 2005 PPP presented in the main text (table 3). Our initial year is 2011, based on the 2011 PPP exchange rates, for which our Gini index is 64.2 and Theil-L index is 83.0. Thus, we focus the comparison on relative changes. G. Ferreira et. al. (2015) describe the differences across several cross-national inequality databases.

<sup>13</sup> Based on Theil-L measure, about 58% of global the inequality is explained by the between-country component in the starting year (2011), compared to a corresponding starting-year figure of 72.9% in the first wave.

|                               |       |       |       |       |
|-------------------------------|-------|-------|-------|-------|
| Percentile 75 / Percentile 25 | 5.4   | 5.2   | 6.5   | 6.6   |
| Mean, \$(PPP)                 | 427.8 | 423.7 | 896.9 | 884.3 |
| Coeff. of variation           | 2.0   | 2.1   | 1.8   | 1.8   |

Source: Ahmed et al. (2016).

Regarding demographic changes, the results suggest that if we consider changes in population, by reweighting the household surveys in order to match the population characteristics predicted for 2030, keeping prices (wages and income) constant, there would not be significant differences on the Gini index for global inequality, while the Theil-L index would be slightly larger. Yet, if we analyze the decomposition between and within countries, our simulations suggest that demographic changes would benefit reductions in inequality between countries, but not within them. We look at the impact of education wave on within-country inequality further when we compare it with the counterfactual case of no education in the next subsection below.

## 4.2 Antimicrobial resistance (AMR)

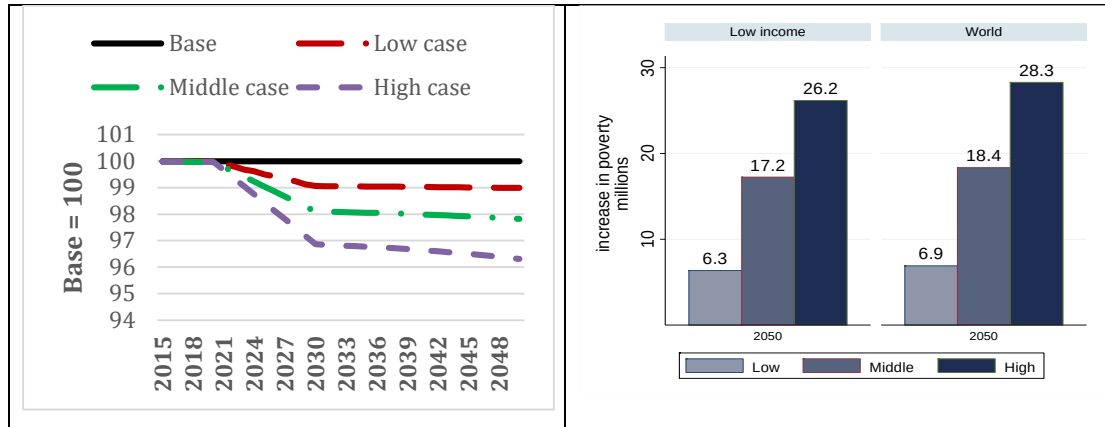
In the second illustration, the report World Bank (2016) uses expert assessments of key factors to form low-, middle-, and high case scenarios for the severity of the impact of AMR, including the deterioration of human health and labor productivity, increases in health care costs; declines in livestock production; and restrictions on global food trade arising from fear.

Relative to a world without AMR, the losses during the period 2015-2050 may add up to 85 trillion dollars in GDP and 23 trillion dollars in global trade.<sup>14 15</sup> Households may have to pay 17 trillion dollars in additional taxes if the rise in healthcare cost is tax-financed. By 2050, an additional 28 million could fall into extreme poverty (\$1.90/day) as a result of AMR (high case). Relative to a \$3.10/day poverty line, the corresponding number is higher at 53 million people. Developing countries will be hurt the most, particularly those with the lowest incomes.

|                          |                                                   |
|--------------------------|---------------------------------------------------|
| Figure 5: Real World GDP | Figure 6: Extreme poverty (PPP\$1.90/day) by 2050 |
|--------------------------|---------------------------------------------------|

<sup>14</sup> A world without AMR refers the baseline projections with zero percent AMR resistance rate (see Appendix 1 for explanation).

<sup>15</sup> In this analysis, all dollar amounts are 2007 US dollars.



Source: World Bank (2016).

## 5 Conclusions

In this paper, we introduce a new database about the skill composition of labor as part of the effort to improve the economic modeling of long-term human capital formation and its effects on poverty and income distribution across countries. Skill definition and accumulation are based on education level as well as demographic characteristics of individuals, which are derived from harmonized household surveys of 104 countries and constructed to provide a consistent definition of labor skills, wages and skill premia. Using the level of schooling to define labor skills is useful for policy analysis, and by itself, the database supplies scarce information about labor to improve the value added of labor in national accounts (such as the SAMs used in CGE modeling).

In particular, the database was designed to improve the global microsimulation model at the World Bank called the GIDD, and we explain how the new features work as part of its documentation and possibly as a template for similar economic modeling. As illustrations, we summarize two cases of the many applications of the framework already made – the first case deals with the global poverty and inequality implications of the coming wave of educated workers from the developing countries, and the second case examines the future threats of the antimicrobial resistance.

[Suggestions for improvement and further research]

## Bibliography

- Ahmed, Amer S., Mauricio Bussolo, Marcio Cruz, Delfin S. Go, and Israel Osorio-Rodarte. 2017. "The Coming Wave of Educated Workers from Developing Countries and Its Potential Impact on Global Inequality." Draft. The World Bank Development Prospects Group.
- Ahmed, Amer S. and Marcio Cruz. 2016a.
- Ahmed, Amer S., Marcio Cruz, Delfin S. Go, Maryla Maliszewska, and Israel Osorio-Rodarte. 2016b. "How Significant Is Africa's Demographic Dividend for Its Future Growth and Poverty Reduction?" *Review of Development Economics*. DOI:10.1111/rode.12227 (online version).
- Balistrèti, Edward J., Maryla Maliszewska, Israel Osorio-Rodarte, David G. Tarr, and Hidemichi Yonezawa. 2016. "Poverty and Shared Prosperity of Deep Integration in Eastern and Southern Africa." World Bank Policy Research Working Paper No. 7660.
- Barro, Robert J. (1997). *Determinants of Economic Growth - A Cross-Country Empirical Study*. The Lionel Robbins Lectures. Cambridge: The MIT Press.
- Bourguignon, François, and Maurizio Bussolo. 2013. "Income Distribution in Computable General Equilibrium Modeling." CHAP. In, 1:1383–1437. *Handbook of Computable General Equilibrium Modeling*. Elsevier.  
<http://ideas.repec.org/h/eee/hacchp/v1y2013icp1383-1437.html>.
- Bourguignon, François, Maurizio Bussolo, and Luiz A Pereira da Silva. 2008. *The Impact of Macroeconomic Policies on Poverty and Income Distribution: Macro-Micro Evaluation Techniques and Tools*. Book. Houndmills, Basingstoke, Hampshire; New York-Washington DC: Palgrave Macmillan; World Bank.  
<http://www.loc.gov/catdir/toc/ecip081/2007040478.html>.
- Bourguignon, François, Francisco H. G. Ferreira, and Phillippe G. Leite. 2008. "Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions." *The Journal of Economic Inequality* 6 (2): 117–48. doi:10.1007/s10888-007-9063-y.
- Bourguignon, François and Luiz Pereira da Silva. 2003. *The Impact of Economic Policies on Poverty and Income Distribution – Evaluation Techniques and Tools*. Oxford University Press.
- Bussolo, Maurizio, Rafael E De Hoyos, and Denis Medvedev. 2010. "Economic Growth and Income Distribution: Linking Macro-Economic Models with Household Survey Data at the Global Level." Article. *International Journal of Microsimulation* 3 (1): 92–103.  
<http://ideas.repec.org/a/ijm/journal/v3y2010i1p92-103.html>.
- Davies, James B. 2009. "Combining Microsimulation with CGE and Macro Modelling for Distributional Analysis in Developing and Transition Countries." Article. *International Journal of Microsimulation* 2 (1): 49–65.
- Deaton, Angus. 2010. "Measuring development: different data, different conclusions?" In *Measure for Measure: How Well Do We Measure Development?* Proceedings of the 8th AFD-EUDN Conference, Paris.
- Deaton, Angus and Valerie Kozel. 2005. *The Great Indian Poverty Debate*. Delhi: Macmillan India Ltd.
- Devarajan, Shantayanan, Delfin S. Go, Maryla Maliszewska, Israel Osorio-Rodarte, and Hans

- Timmer. 2015. "Stress-Testing Africa's Recent Growth and Poverty Performance." *Journal of Policy Modeling* 37 (4): 521–47. doi:10.1016/j.jpolmod.2015.04.006.
- Hall, Robert E. and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114 (February): 83-116.
- Hendricks, Lutz (2002). "How Important is Human Capital for Development? Evidence from Immigrant Earnings." *American Economic Review* 92 (March): 198-219.
- Ianchovichina, Elena I. and Terrie L. Walmsley, eds. (2012). *Dynamic Modeling and Applications for Global Economic Analysis*. Cambridge: Cambridge University Press.
- Klenow, Peter J. and Andrés Rodríguez-Clare (1997). "The Neoclassical Revival in Growth Economics: Has it Gone Too Far?" *NBER Macroeconomics Annual* 12: 73-103.
- Lakatos, Csilla, Maryla Maliszewska, Israel Osorio-Rodarte, and Delfin Sia Go. 2016. "China's Slowdown and Rebalancing: Potential Growth and Poverty Impacts on Sub-Saharan Africa." World Bank Policy Research Working Paper No. 7666.
- McDonald, S., K. Thierfelder, and S. Robinson (2007). "GLOBE: A SAM-based global CGE model using GTAP data," Economics Working Paper, US Naval Academy, Annapolis, USA.
- Montenegro, Claudio E. and Maximilian L. Hirn. (2009). "A New Disaggregated Set of Labor Market Indicators Using Standardized Household Surveys from Around the World." Background paper for the 2009 World Development Report, *Reshaping Economic Geography*. Washington, DC: World Bank.  
<https://openknowledge.worldbank.org/handle/10986/9033>.
- Narayanan, B., Aguiar, A. & McDougall, R., 2012. *Global Trade, Assistance, and Production: The GTAP 8 Data Base*, Center for Global Trade Analysis: Purdue University, West Lafayette IN.
- Tyers, Rod and Qun Shi (2012). "Global Demographic Change, Labor Force Growth, and Economic Performance." In Elena I. Ianchovichina and Terrie L. Walmsley, eds. *Dynamic Modeling and Applications for Global Economic Analysis*. Cambridge: Cambridge University Press, pp. 342-75.
- United Nations, World Population Prospects: The 2012 Revision, DVD Edition, New York: United Nations, Department of Economic and Social Affairs, Population Division (2013).
- van der Mensbrugghe, Dominique. 2011. "LINKAGE Technical Reference Document Version 7.1." [http://siteresources.worldbank.org/INTPROSPECTS/Resources/334934-1314986341738/TechRef7.1\\_01Mar2011.pdf](http://siteresources.worldbank.org/INTPROSPECTS/Resources/334934-1314986341738/TechRef7.1_01Mar2011.pdf).
- Wittenberg, Martin. 2010. "An Introduction to Maximum Entropy and Minimum Cross-Entropy Estimation Using Stata." *The Stata Journal* 10 (3): 315–30.
- World Bank (2014). *The Economic Impact of the 2014 Ebola Epidemic – Short and Medium Term Estimates for West Africa*. A World Bank Report.
- World Bank (2016). *Drug-Resistant Infections: A Threat to Our Economic Future*. A World Bank Discussion Draft and publication.

## Appendix I: Definitions and countries in the GIDD-SWB

Table AI.1: The variables used to generate the GIDD-SWB from the I2D2

---

|   |                                                         |
|---|---------------------------------------------------------|
| • | <i>age</i> : Age (individual)                           |
| • | <i>gender</i> : Gender (individual)                     |
| ○ | <i>gender</i> = 1 "Male"                                |
| ○ | <i>gender</i> = 2 "Female"                              |
| • | <i>educy</i> : Years of schooling                       |
| • | <i>lstatus</i> : Labor status (individual)              |
| ○ | <i>lstatus</i> = 1 "Employed"                           |
| ○ | <i>lstatus</i> = 2 "Unemployed"                         |
| ○ | <i>lstatus</i> = 3 "Not part of labor force"            |
| • | <i>empstat</i> : Employment status (individual)         |
| ○ | <i>empstat</i> = 1 "Paid employee"                      |
| ○ | <i>empstat</i> = 2 "Non-paid employee"                  |
| ○ | <i>empstat</i> = 3 "Employer"                           |
| ○ | <i>empstat</i> = 4 "Self-employed"                      |
| • | <i>wage</i> : Last wage payment (individual)            |
| • | <i>unitwage</i> : Last wage's time unit (individual)    |
| ○ | <i>unitwage</i> = 1 "Daily"                             |
| ○ | <i>unitwage</i> = 2 "Weekly"                            |
| ○ | <i>unitwage</i> = 3 "Every two weeks"                   |
| ○ | <i>unitwage</i> = 4 "Bimonthly"                         |
| ○ | <i>unitwage</i> = 5 "Monthly"                           |
| ○ | <i>unitwage</i> = 6 "Trimester"                         |
| ○ | <i>unitwage</i> = 7 "Biannual"                          |
| ○ | <i>unitwage</i> = 8 "Annually"                          |
| ○ | <i>unitwage</i> = 9 "Hourly"                            |
| ○ | <i>unitwage</i> = 10 "Other"                            |
| • | <i>pci</i> : Monthly income per capita (household)      |
| • | <i>pcc</i> : Monthly consumption per capita (household) |
| • | <i>industry</i> : <i>Industry</i> (individual)          |
| ○ | <i>industry</i> = 1 "Agriculture, Hunting, etc."        |
| ○ | <i>industry</i> = 2 "Mining"                            |
| ○ | <i>industry</i> = 3 "Manufacturing"                     |
| ○ | <i>industry</i> = 4 "Public services"                   |
| ○ | <i>industry</i> = 5 "Construction"                      |
| ○ | <i>industry</i> = 6 "Retail and Hotels"                 |
| ○ | <i>industry</i> = 7 "Transport and Telecommunications"  |
| ○ | <i>industry</i> = 8 "Finance and business services"     |
| ○ | <i>industry</i> = 9 "Communal services"                 |
| ○ | <i>industry</i> = 10 "Other"                            |

---

**Table AI.2 I2D2 industry classification vis-à-vis the ISIC definition**

| I2D2 |                               | ISIC   |                                                      |
|------|-------------------------------|--------|------------------------------------------------------|
| Code | Description                   | Code   | Description                                          |
| 01   | Agriculture, Hunting, etc.    | 01, 02 | Agriculture, hunting and forestry                    |
|      |                               | 05     | Fishing                                              |
| 02   | Mining                        | 10-14  | Mining and quarrying                                 |
| 03   | Manufacturing                 | 15-37  | Manufacturing                                        |
| 04   | Public services               | 40,41  | Electricity, gas and water supply                    |
| 05   | Construction                  | 45     | Construction                                         |
| 06   | Retail, Hotels                | 50-52  | Wholesale and retail trade vehicles, repair of motor |
|      |                               | 55     | Hotels and restaurants                               |
| 07   | Transport and telecom.        | 60-64  | Transport, storage and communications                |
| 08   | Finance and business services | 65-67  | Financial intermediation                             |
|      |                               | 70-74  | Real state, rentin and business activity             |
| 09   | Communal Services             | 75     | Public administration                                |
|      |                               | 80     | Education                                            |
|      |                               | 85     | Health and social work                               |
|      |                               | 90-93  | Other community, social and personal services        |
|      |                               | 95     | Activities of private households as employers        |
| 10   | Others                        | 99     | Extra-territorial organizations and bodies           |

Source: Montegro and Him (2009).

**Table AI.3 List of countries, type of survey, and survey year in GIDD Sectoral Wage Bill**

| Country | Year | Country name          | Survey name                             | Type   |
|---------|------|-----------------------|-----------------------------------------|--------|
| CHN     | 2002 |                       |                                         |        |
| FJI     | 2008 | Fiji                  | Household Income and Expenditure Survey | HIES   |
| FSM     | 2000 | Micronesia, Fed. Sts. | Census                                  | CENSUS |
| IDN     | 2007 | Indonesia             | SUSENAS                                 | LSS    |
| KHM     | 2007 | Cambodia              | Cambodia Socio-Economic Survey          | LSS    |
| LAO     | 2008 | Lao PDR               | Lao Expenditure and Consumption Survey  | HIES   |
| MNG     | 2009 | Mongolia              | Socio-Economic Survey                   | LSS    |
| MYS     | 2010 | Malaysia              | Labour Force Survey                     | LFS    |
| PHL     | 2007 | Philippines           | Labour Force Survey                     | LFS    |
| PNG     | 2010 | Papua New Guinea      | Household and Income Expenditure Survey | HIES   |
| SLB     | 2005 | Solomon Islands       | Household Income and Expenditure Survey | HIES   |

|     |      |                        |                                                              |      |
|-----|------|------------------------|--------------------------------------------------------------|------|
| THA | 2006 | Thailand               | Household Income and Expenditure Survey                      | LFS  |
| TMP | 2007 |                        | Survey of Living Standards                                   | LSS  |
| TON | 1996 | Tonga                  | Household Income and Expenditure Survey                      | HIES |
| VNM | 2004 | Vietnam                |                                                              |      |
| ALB | 2005 | Albania                | Living Standard Measurement Survey                           | LSMS |
| ARM | 1999 | Armenia                | Integrated Survey of Living Standards                        | LSS  |
| AZE | 1995 | Azerbaijan             | Standard of Living in Azerbaijan                             | LSS  |
| BGR | 2007 | Bulgaria               | EUROSILC                                                     | LSS  |
| BIH | 2001 | Bosnia and Herzegovina | Living Standards Measurement Survey                          | LSS  |
| BLR | 1998 | Belarus                | Integrated Household Survey                                  | HIES |
| GEO | 2010 | Georgia                | Monitoring of Households                                     | HIES |
| HUN | 2007 | Hungary                | EUROSILC                                                     | LSS  |
| KGZ | 1997 | Kyrgyz Republic        | Household Budget Survey                                      | HIES |
| KSV | 2003 | Kosovo                 | Labour Force Survey                                          | LFS  |
| MDA | 2005 | Moldova                | Integrated Household Survey                                  | HIES |
| MKD | 2005 | Macedonia, FYR         | Labour Force Survey                                          | LFS  |
| ROM | 2007 |                        | EUROSILC                                                     | LSS  |
| SRB | 2008 | Serbia                 | Labour Force Survey                                          | LFS  |
| TJK | 2003 | Tajikistan             | Tajikistan Living Standard Survey                            | LSS  |
| TKM | 1998 | Turkmenistan           | Living Standard Measurement Survey                           | LSMS |
| TUR | 2007 | Turkey                 | Labor Force Survey                                           | LFS  |
| UKR | 2007 |                        |                                                              |      |
| ARG | 2007 | Argentina              | Encuesta Permanente de Hogares                               | LSS  |
| BLZ | 1999 | Belize                 | Labor Force Survey                                           | LSS  |
| BOL | 2007 | Bolivia                | Encuesta Continua de Hogares                                 | LSS  |
| BRA | 2007 | Brazil                 | Pesquisa Nacional por Amostra de Domicílios                  | LSS  |
| COL | 2007 | Colombia               | Encuesta Continua de Hogares                                 | LSS  |
| CRI | 2007 | Costa Rica             | Encuesta de Hogares de Propósitos Múltiples                  | LSS  |
| DOM | 2007 | Dominican Republic     | Encuesta Nacional de Fuerza de Trabajo                       | LSS  |
| ECU | 2007 | Ecuador                | Encuesta de Empleo, Desempleo y Subempleo                    | LSS  |
| GTM | 2006 | Guatemala              | Encuesta Nacional de Condiciones de Vida                     | LSS  |
| GUY | 1992 | Guyana                 | Household Income and Expenditure Survey-Living               | LSS  |
| HND | 2007 | Honduras               | Encuesta Permanente de Hogares de Propósitos Múltiples       | LSS  |
| HTI | 2001 | Haiti                  | Enquête sur les Conditions de Vie en Haïti                   | LSS  |
| JAM | 2002 | Jamaica                | Jamaica Survey of Living Conditions                          | LSS  |
| MEX | 2008 | Mexico                 | Encuesta Nacional de Ingresos y Gastos de los Hogares        | LSS  |
| NIC | 2005 | Nicaragua              | Encuesta Nacional de Hogares sobre Medición de Nivel de Vida | LSS  |
| PAN | 2006 | Panama                 | Encuesta de Hogares                                          | LSS  |
| PER | 2007 | Peru                   | Encuesta Nacional de Hogares                                 | LSS  |
| PRY | 2007 | Paraguay               | Encuesta Permanente de Hogares                               | LSS  |
| SLV | 2007 | El Salvador            | Encuesta de Hogares de Propósitos Múltiples                  | LSS  |
| SUR | 1999 | Suriname               | Expenditure Household Survey                                 | LSS  |
| VEN | 2006 | Venezuela, RB          | Encuesta de Hogares Por Muestreo                             | LSS  |
| DJI | 2002 | Djibouti               | Echantillon Maitre                                           | LSS  |
| EGY | 2006 | Egypt, Arab Rep.       | Egyptian Labour Market Survey                                | LFS  |
| IRQ | 2006 | Iraq                   | Iraq Household Socio Economic Survey                         | LSS  |

| JOR     | 2002 | Jordan                | Household Income and Expenditure Survey                         | HIES |
|---------|------|-----------------------|-----------------------------------------------------------------|------|
| MAR     | 1998 | Morocco               |                                                                 |      |
| SYR     | 2004 | Syrian Arab Republic  | Family Income and Expenditure                                   | HIES |
| TUN     | 2001 | Tunisia               | ENQUETE NATIONALE SUR LA POPULATION ET L'EMPLOI DE 2001         | LFS  |
| YEM     | 2005 | Yemen, Rep.           | Household Budget Survey                                         | LFS  |
| AFG     | 2007 | Afghanistan           | National Risk and Vulnerability Assessment                      | LSS  |
| BGD     | 2005 | Bangladesh            | Household Income and Expenditure Survey                         | HIES |
| BTN     | 2007 | Bhutan                | Bhutan Living Standards Survey                                  | LSS  |
| IND     | 2007 | India                 | Socio-Economic Survey                                           | LSS  |
| LKA     | 2008 | Sri Lanka             | Labour Force Survey                                             | LFS  |
| MDV     | 2004 | Maldives              | Vulnerability and poverty assessment survey                     | LSS  |
| NPL     | 2008 | Nepal                 | Labour Force Survey                                             | LFS  |
| PAK     | 2007 | Pakistan              | Labour Force Survey                                             | LFS  |
| Country | Year | Country name          | Survey name                                                     | Type |
| AGO     | 1999 | Angola                | Inquerito aos Agredados Familiares Sobre Despesas E Receitas    | HIES |
| BFA     | 2010 | Burkina Faso          |                                                                 |      |
| BWA     | 2009 | Botswana              | Botswana Core Welfare Indicators (Poverty) Survey               | LSS  |
| CIV     | 2008 | Cote d'Ivoire         |                                                                 |      |
| CMR     | 2007 | Cameroon              | ECAM3                                                           | LSS  |
| COM     | 2004 | Comoros               | Enquete Integral Aupres des Menages                             | LSS  |
| CPV     | 2007 | Cape Verde            | Questionário Unificado de Indicadores Básicos de Bem-Estar      | LSS  |
| ETH     | 2000 | Ethiopia              | Household Income and Expenditure                                | HIES |
| GAB     | 2005 | Gabon                 | Enquete Gabonaise Pour l'Evaluation Et Le Suivi De La Pauvrete  | LSS  |
| GHA     | 2005 | Ghana                 | GHANA LIVING STANDARDS SURVEY ROUND THREE (GLSS 5)              | LSS  |
| GIN     | 1994 | Guinea                | Enquete Integrale Sur Les Conditions de Vie des Menages         | LSS  |
| GMB     | 1998 | Gambia, The           | Household Poverty Survey                                        | LSS  |
| KEN     | 2005 | Kenya                 | Kenya Integrated Budget Survey                                  | LSS  |
| MDG     | 2010 | Madagascar            |                                                                 |      |
| MLI     | 1994 | Mali                  | Enquete Malienne De Conjoncture Economique et Sociale           | LSS  |
| MOZ     | 2008 | Mozambique            | HOUSEHOLD BUDGET SURVEY - IOF                                   | HIES |
| MRT     | 2000 | Mauritania            | ENQUETE PERMANENTE SUR LES CONDITIONS DE VIE DES MENAGES        | LSS  |
| MUS     | 2007 | Mauritius             | Continuous Multi-purpose Household Survey                       | LSS  |
| MWI     | 2010 |                       |                                                                 |      |
| NAM     | 1993 | Namibia               | Household Income and Expenditure Survey                         | LFS  |
| NER     | 2007 | Niger                 |                                                                 |      |
| NGA     | 2003 | Nigeria               | Nigeria Living Standards Survey                                 | LSS  |
| RWA     | 2005 | Rwanda                | Enquete Integrale Sur Les Conditions de Vie des Menages         | LSS  |
| SEN     | 2001 | Senegal               | Enquete Senegalaise Aupres Des Menages (EASM II)                | LSS  |
| SLE     | 2003 | Sierra Leone          | SIERRA LEONE INTEGRATED HOUSEHOLD SURVEY (SLIHS)                | LSS  |
| STP     | 2000 | Sao Tome and Principe | Enquete Sur les Conditions des Menages                          | LSS  |
| SWZ     | 2000 | Swaziland             | Household Income and Expenditure Survey                         | HIES |
| TCD     | 2003 | Chad                  | 2nd ENQUÊTE SUR LA CONSOMMATION ET LE SECTEUR INFORMEL AU TCHAD | HIES |
| TGO     | 2006 | Togo                  | Questionnaire des Indicateurs Base du Bien-être (QUIBB)         | LSS  |
| TZA     | 2006 | Tanzania              | INTEGRATED LABOUR FORCE SURVEY                                  | LSS  |
| UGA     | 2005 | Uganda                | Uganda National Household Survey                                | LSS  |
| ZAF     | 2007 | South Africa          | Labor Force Survey                                              | LFS  |

|     |      |                 |                                                     |        |
|-----|------|-----------------|-----------------------------------------------------|--------|
| ZAR | 2005 |                 | ENQUETE NATIONALE DU TYPE 1 2 3 AUPRES DES MENAGES  | LSS    |
| ZMB | 2010 | Zambia          |                                                     |        |
| AUS | 2007 | Australia       |                                                     |        |
| AUT | 2007 | Austria         | EUROSILC                                            | LSS    |
| BEL | 2007 | Belgium         | EUROSILC                                            | LSS    |
| BHS | 2001 | Bahamas, The    | Living Conditions Survey                            | LSS    |
| BRB | 1996 | Barbados        | Continuous Labour Force Sample Survey               | LFS    |
| CAN | 2001 | Canada          | Census                                              | CENSUS |
| CHL | 2006 | Chile           | Encuesta de Caracterización Socioeconómica Nacional | LSS    |
| CYP | 2007 | Cyprus          | EUROSILC                                            | LSS    |
| CZE | 2007 | Czech Republic  | EUROSILC                                            | LSS    |
| DEU | 2007 | Germany         | EUROSILC                                            | LSS    |
| DNK | 2007 | Denmark         | EUROSILC                                            | LSS    |
| ESP | 2007 | Spain           | EUROSILC                                            | LSS    |
| EST | 2007 | Estonia         | EUROSILC                                            | LSS    |
| FIN | 2007 | Finland         | EUROSILC                                            | LSS    |
| FRA | 2007 | France          | EUROSILC                                            | LSS    |
| GBR | 2007 | United Kingdom  | EUROSILC                                            | LSS    |
| GRC | 2007 | Greece          | EUROSILC                                            | LSS    |
| HRV | 2004 | Croatia         | Labour Force Survey                                 | LFS    |
| IRL | 2007 | Ireland         | EUROSILC                                            | LSS    |
| ISL | 2007 | Iceland         | EUROSILC                                            | LSS    |
| ITA | 2007 | Italy           | EUROSILC                                            | LSS    |
| LTU | 2007 | Lithuania       | EUROSILC                                            | LSS    |
| LUX | 2007 | Luxembourg      | EUROSILC                                            | LSS    |
| LVA | 2007 | Latvia          | EUROSILC                                            | LSS    |
| MLT | 2009 | Malta           | EUROSILC                                            | LSS    |
| NLD | 2007 | Netherlands     | EUROSILC                                            | LSS    |
| NOR | 2007 | Norway          | EUROSILC                                            | LSS    |
| POL | 2007 | Poland          | EUROSILC                                            | LSS    |
| PRI | 2005 | Puerto Rico     | Census                                              | CENSUS |
| PRT | 2007 | Portugal        | EUROSILC                                            | LSS    |
| RUS | 2007 |                 |                                                     |        |
| SVK | 2007 | Slovak Republic | EUROSILC                                            | LSS    |
| SVN | 2007 | Slovenia        | EUROSILC                                            | LSS    |
| SWE | 2007 | Sweden          | EUROSILC                                            | LSS    |
| URY | 2007 | Uruguay         | Encuesta Continua de Hogares                        | LSS    |

## Appendix II: Number of workers and average wage in the GIDD-SWB

**Table AII.1 The Number of Paid Workers by Industry and Income Level (in 1,000)**

| Skill level | Industry                        | High-Income<br>OECD | High-Income<br>Non-OECD | Low Income | Lower<br>Middle | Upper-<br>Middle | Total     |
|-------------|---------------------------------|---------------------|-------------------------|------------|-----------------|------------------|-----------|
| Unskilled   | Agriculture                     | 3,421               | 4,862                   | 73,659     | 196,336         | 308,751          | 587,029   |
|             | Mining                          | 88                  | 15,192                  | 1,042      | 4,324           | 10,087           | 30,733    |
|             | Manufacturing                   | 9,757               | 32,183                  | 9,845      | 44,082          | 107,168          | 203,035   |
|             | Commerce                        | 10,111              | 274                     | 13,908     | 46,151          | 65,859           | 136,304   |
|             | Community<br>services           | 6,118               | 174                     | 1,954      | 12,144          | 40,771           | 61,160    |
|             | Construction                    | 6,307               | 188                     | 4,393      | 42,633          | 51,032           | 104,552   |
|             | Financial and<br>business       | 3,156               | 37                      | 882        | 2,286           | 8,108            | 14,470    |
|             | Others                          | 3,726               | 142                     | 4,621      | 14,223          | 57,959           | 80,671    |
|             | Public utilities                | 28                  | 6                       | 381        | 1,057           | 3,760            | 5,232     |
|             | Transports and<br>communication | 2,667               | 68                      | 5,022      | 21,139          | 20,946           | 49,841    |
| Skilled     | Agriculture                     | 5,576               | 1,453                   | 4,292      | 34,376          | 27,773           | 73,469    |
|             | Mining                          | 1,073               | 6,655                   | 235        | 5,976           | 3,335            | 17,274    |
|             | Manufacturing                   | 49,054              | 21,273                  | 2,197      | 36,656          | 65,464           | 174,643   |
|             | Commerce                        | 56,025              | 889                     | 3,375      | 32,161          | 47,865           | 140,315   |
|             | Community<br>services           | 83,845              | 1,260                   | 3,011      | 27,221          | 84,939           | 200,276   |
|             | Construction                    | 20,694              | 314                     | 687        | 9,948           | 13,820           | 45,463    |
|             | Financial and<br>business       | 46,474              | 335                     | 1,194      | 11,841          | 26,112           | 85,957    |
|             | Others                          | 10,333              | 234                     | 5,155      | 20,010          | 37,119           | 72,853    |
|             | Public utilities                | 1,314               | 19                      | 261        | 1,734           | 6,584            | 9,912     |
|             | Transports and<br>communication | 16,703              | 296                     | 1,069      | 12,273          | 20,313           | 50,653    |
| Total       |                                 | 336,469             | 85,855                  | 137,184    | 576,569         | 1,007,764        | 2,143,841 |

Source: GIDD Sectoral Wage Bill Database

Note: HIO: High-Income countries (OECD); HIX: High-Income Countries (non-OECD); LIC: Low-Income Countries; LMI: Lower Middle-Income; UMI: Upper Middle-Income.

**Table AII.2 Average Wages by Skill Level and Industry**

| Skill level | Industry                 | Wages, PPP\$(2007) |       |     |     |     | Wages, US\$(2007) |       |     |     |     |
|-------------|--------------------------|--------------------|-------|-----|-----|-----|-------------------|-------|-----|-----|-----|
|             |                          | HIO                | HIX   | LIC | LMI | UMI | HIO               | HIX   | LIC | LMI | UMI |
| Unskilled   | Agriculture              | 956                | 516   | 128 | 136 | 118 | 1,086             | 316   | 82  | 58  | 72  |
|             | Mining                   | 2,818              | 572   | 184 | 155 | 399 | 2,381             | 346   | 125 | 74  | 261 |
|             | Manufacturing            | 1,509              | 575   | 148 | 170 | 254 | 1,792             | 349   | 66  | 79  | 169 |
|             | Public utilities         | 2,536              | 732   | 138 | 265 | 316 | 2,021             | 588   | 60  | 102 | 199 |
|             | Construction             | 1,583              | 582   | 178 | 181 | 208 | 1,767             | 742   | 70  | 79  | 135 |
|             | Commerce                 | 1,229              | 419   | 183 | 183 | 255 | 1,436             | 498   | 123 | 94  | 178 |
|             | Transports and communic. | 1,735              | 716   | 145 | 209 | 313 | 2,095             | 795   | 69  | 97  | 219 |
|             | Financial and business   | 1,556              | 383   | 117 | 214 | 347 | 1,768             | 365   | 59  | 108 | 243 |
|             | Community services       | 1,353              | 487   | 200 | 236 | 243 | 1,660             | 550   | 97  | 129 | 170 |
|             | Others                   | 904                | 228   | 121 | 155 | 138 | 1,153             | 237   | 46  | 65  | 91  |
| Skilled     | Agriculture              | 2,165              | 695   | 184 | 212 | 183 | 2,263             | 430   | 91  | 87  | 117 |
|             | Mining                   | 5,654              | 786   | 224 | 463 | 658 | 4,740             | 475   | 148 | 191 | 431 |
|             | Manufacturing            | 3,841              | 801   | 273 | 392 | 441 | 3,664             | 500   | 116 | 160 | 309 |
|             | Public utilities         | 6,386              | 1,260 | 317 | 394 | 465 | 5,092             | 2,222 | 126 | 158 | 341 |
|             | Construction             | 3,291              | 756   | 294 | 288 | 456 | 3,124             | 838   | 115 | 116 | 310 |
|             | Commerce                 | 2,451              | 636   | 253 | 274 | 494 | 2,298             | 818   | 139 | 127 | 335 |
|             | Transports and communic. | 3,282              | 858   | 305 | 332 | 482 | 3,221             | 952   | 138 | 136 | 336 |
|             | Financial and business   | 4,332              | 1,273 | 390 | 377 | 721 | 4,073             | 1,760 | 151 | 164 | 497 |
|             | Community services       | 3,311              | 815   | 398 | 430 | 559 | 3,252             | 1,131 | 141 | 188 | 386 |
|             | Others                   | 1,564              | 729   | 266 | 249 | 303 | 2,000             | 1,112 | 97  | 108 | 234 |

Source: GIDD Sectoral Wage Bill Database

## Appendix III: Symbols and equations in GIDD 2.0

### Variables and symbols

Table AIII.1 defines the set of LAVs employed in a typical GIDD application. For simplicity, we will omit the geographical dimension  $c$  for the rest of the paper. Nomenclature listed in the third column (from the left to right) defines symbols used in the equations below. Items in the column marked as “micro application” (the last column to the right) list values used in the empirical applications or illustrations. In the GIDD, dynamics are driven strongly by changes in the demographic structure  $\{m_t, g_t, s_t\}$ . For most applications, demographic changes are exogenously determined in both, the macro and micro models. An important distinction is that employed individuals can receive both labor and non-labor income. The unemployed and all individuals not considered part of the labor force can only be recipients of non-labor income. In other words, the unemployed and individuals not in the labor force must have labor incomes equal to zero.

Table AIII.2 describes a basic classification of individuals for the household survey, with each individual belonging to one of the mutually exclusive groups. Regarding the status labor force participation status, the unemployed and individuals not in the labor force can conveniently be assigned to the *null* economic sector  $p_0$ , while employed individuals can be placed in the shrinking,  $p_1$ , or expanding sector,  $p_2$ . Lastly, labor is defined as skilled (more educationally qualified) and unskilled (less educationally qualified).

Typically, household surveys would also permit to classify employed individuals, according to their main job, as unpaid workers, paid workers, the self-employed, and employers. For the case of the self-employed and employers, labor income would ideally represent the marginal contribution from labor to total output. Capital gains, in this respect, should ideally be considered part of non-labor income. Empirically, the quality and scope of harmonized survey data usually limits the analysis and force the microsimulation to rely on a set of assumptions.

Table AIII.3 maps the set of macroeconomic projections from a typical scenario to the household variables in GIDD. A typical macro scenario file used for simulation is available for download on the World Bank’s GIDD website.

Table AIII.1 Basic LAV structure of the GIDD model

| Item  | Definition                 | Nomenclature                        | Micro application                         |
|-------|----------------------------|-------------------------------------|-------------------------------------------|
| $t$   | {initial, counterfactuals} | $\{t_0, t_1\}$                      | {2010, 2025}                              |
| $m_t$ | {8 age groups}             | $\{m_{1t}, m_{2t}, \dots, m_{8t}\}$ | {0-9, 10-19, ..., 50-59, 60+}             |
| $g_t$ | {gender}                   | $\{g_{1t}, g_{2t}\}$                | {female, male}                            |
| $s_t$ | {3 education levels}       | $\{s_{1t}, s_{2t}, s_{3t}\}$        | {0-6, 7-12, 13+}                          |
| $f$   | {2 labor status}           | $\{f_0, f_1\}$                      | {unemployed/not in labor force, employed} |
| $p$   | {3 sectors}                | $\{p_0, p_1, p_2\}$                 | {null, shrinking, expanding}              |
| $q$   | {2 types of workers}       | $\{q_1, q_2\}$                      | {unskilled, skilled}                      |
| $b$   | {2 bundle of goods}        | $\{b_1, b_2\}$                      | {food, non-food}                          |

Source: Author's own elaboration

Table AIII.2 GIDD: Basic classification of all individuals in the household survey

| by labor status $f$                       | by sector of employment $p$ | by qualification $q$ | with labor and non-labor incomes                         |
|-------------------------------------------|-----------------------------|----------------------|----------------------------------------------------------|
| Unemployed /<br>Not in the labor<br>force | None                        | Unskilled            | $z_{it_0 p_0 q_1} = 0;$<br>$w_{it_0 p_0 q_1} \geq 0$     |
|                                           |                             | Skilled              | $z_{it_0 p_0 q_2} = 0;$<br>$w_{it_0 p_0 q_2} \geq 0$     |
| Employed                                  | Shrinking sector            | Unskilled            | $z_{it_0 p_1 q_0} \geq 0;$<br>$w_{it_0 p_1 q_0} \geq 0;$ |
|                                           |                             | Skilled              | $z_{it_0 p_1 q_1} \geq 0;$<br>$w_{it_0 p_1 q_1} \geq 0$  |
|                                           | Expanding sector            | Unskilled            | $z_{it_0 p_2 q_0} \geq 0;$<br>$w_{it_0 p_2 q_0} \geq 0$  |
|                                           |                             | Skilled              | $z_{it_0 p_2 q_2} \geq 0;$<br>$w_{it_0 p_2 q_2} \geq 0$  |

Note:  $\forall i \in h$ 

Source: Author's own elaboration

Table AIII.3 Definition of variables in the macro scenario and its correspondence with micro aggregates

| Linkage Aggregate Variable | Definition                                    | Equivalence on the household survey                                            |
|----------------------------|-----------------------------------------------|--------------------------------------------------------------------------------|
| l_p1_q0                    | Labor supply in shrinking sector of unskilled | $\sum_i \theta_i \quad \forall i \in p_1 \cap q_0$                             |
| l_p1_q1                    | Labor supply in shrinking sector of skilled   | $\sum_i \theta_i \quad \forall i \in p_1 \cap q_1$                             |
| l_p2_q0                    | Labor supply in expanding sector of unskilled | $\sum_i \theta_i \quad \forall i \in p_2 \cap q_0$                             |
| l_p2_q1                    | Labor supply in expanding sector of skilled   | $\sum_i \theta_i \quad \forall i \in p_2 \cap q_1$                             |
| w_p1_q0                    | Wage index in shrinking sector for unskilled  | $\frac{\sum_i \theta_i w_i}{\sum_i \theta_i} \quad \forall i \in p_1 \cap q_0$ |
| w_p1_q1                    | Wage index in shrinking sector for skilled    | $\frac{\sum_i \theta_i w_i}{\sum_i \theta_i} \quad \forall i \in p_1 \cap q_1$ |
| w_p2_q0                    | Wage index in expanding sector for unskilled  | $\frac{\sum_i \theta_i w_i}{\sum_i \theta_i} \quad \forall i \in p_2 \cap q_0$ |
| w_p2_q1                    | Wage index in expanding sector for skilled    | $\frac{\sum_i \theta_i w_i}{\sum_i \theta_i} \quad \forall i \in p_2 \cap q_1$ |
| I                          | Per capita household income aggregate         | $\frac{\sum_i \theta_i y_i}{\sum_i \theta_i}$                                  |
| P_b1                       | Price index for food                          | 1                                                                              |
| P_b2                       | Price index for non-food                      | 1                                                                              |

Source: Authors' own elaboration

### Setting household sampling weights in the base year

We first define nominal income of individual  $i$  in household  $h$  as the sum of labor,  $w_{iht}$  and non-labor,  $z_{iht}$  income

$$y_{thi} = w_{thi} + z_{thi} \quad \forall i \in h \wedge t = \{t_0, t_1\} \quad (A1)$$

As a result, household total nominal income follows from Equation (A2)

$$Y_{ht} = \sum_{i \in h} y_{iht} \quad (\text{A2})$$

Real household total income is obtained by dividing the previous expression by a price deflator  $P_{ht}(\alpha_h, \bar{P}_{b_0t}, \bar{P}_{b_1t})$ :

$$Y_{ht}^P = \frac{Y_{ht}}{P_{ht}(\alpha_h, \bar{P}_{b_0t}, \bar{P}_{b_1t})} = \frac{Y_{ht}}{\alpha_{ht}\bar{P}_{b_0t} + (1 - \alpha_{ht})\bar{P}_{b_1t}} \quad (\text{A3})$$

where  $\alpha_{ht}$  is the budget share spent on food by household  $h$ . Typically, the initial index of consumer prices is normalized to 1 so that:

$$\bar{P}_{b_0t_0} = \bar{P}_{b_1t_0} = 1 \quad (\text{A4})$$

which results in

$$Y_{ht_0}^P = Y_{ht_0} \quad (\text{A5})$$

In the household survey, total household per capita income  $\bar{I}_{t_0}$  can be obtained from Equation (A6):

$$\bar{I}_{t_0} = \frac{\sum_h \theta_{ht_0} Y_{ht_0}^P}{\sum_h \theta_{ht_0}} \quad (\text{A6})$$

where  $\theta_{ht_0}$  is the household-specific frequency sampling weight.

### Revising the household sampling weights to account for changes in the demographic structure

The next step in the microsimulation is to establish a common set of pathways for the demographic structure  $\{m_t \cdot g_t \cdot s_t\}$  in both the macro and micro models. The alternative demographic pathways are exogenously determined. While the base year values  $\{m_{t_0} \cdot g_{t_0} \cdot s_{t_0}\}$  are obtained directly from the household survey, demographic structures from  $t_0$  to  $t_N$  are constructed from the assumptions of the baseline and counterfactual scenarios, relying on external sources as desired.<sup>16</sup>

Once the demographic structure  $\{m_{t_1} \cdot g_{t_1} \cdot s_{t_1}\}$  has been set, the objective is to obtain a set of household sampling weights that replicate the counterfactual demographic structure using the minimum-cross entropy method proposed by Wittenberg (2010). This method allows the recreation of homogenous weights within household members. In most household surveys, the

---

<sup>16</sup> It is important to highlight that there might exist discrepancies between the  $\{m_{t_0} \cdot g_{t_0} \cdot s_{t_0}\}$  structure found in the microdata and any other initial structure  $\{\hat{m}_{t_0} \cdot \hat{g}_{t_0} \cdot \hat{s}_{t_0}\}$  collected from external sources.

unit of analysis and sampling is the household; hence, sampling weights attached to every individual within the household are equal. It follows that:

$$\sum_h \sum_i \theta_{iht_0} = H_{t_0} \quad (A7)$$

$H_{t_0}$  is the total population in time  $t_0$ , and  $\theta_{iht_0}$  represents the initial frequency weight of individual  $i$  within household  $h$ ,

$$\theta_{iht_0} = \theta_{jht_0} = \theta_{ht_0} \quad \text{for } i \neq j \wedge i, j \in h$$

where  $\theta_{iht_0}$  equal to the common weight  $\theta_{ht_0}$ . Equation (A7) can be rewritten as equation (A8):

$$\sum_h \theta_{ht_0} \cdot \text{household size}_h = \sum_h \theta_{ht_0}^* = H_{t_0} \quad (A8)$$

Rewriting Equation (12) in terms of probabilities:

$$1 = \sum_h \frac{\theta_{ht_0}^*}{H_{t_0}} = \sum_h \vartheta_{ht_0}^* \quad (A9)$$

The new structure of total population to estimate is:

$$\sum_h \sum_i \tilde{\theta}_{iht_1} 1[m_{t_1}, g_{t_1}, s_{t_1}] = \tilde{H}_{t_1} 1[m_{t_1}, g_{t_1}, s_{t_1}] \quad (A10)$$

where  $1[m_{t_1}, g_{t_1}, s_{t_1}]$  is the indicator function,  $\tilde{H}_{t_1}$  is the new total population, and each member  $i$  of household  $h$  has the same weight,  $\tilde{\theta}_{iht_1} = \tilde{\theta}_{jht_1} = \tilde{\theta}_{ht_1}$  for  $i \neq j \wedge i, j \in h$ , then

$$\sum_h \tilde{\theta}_{ht_1} \text{household size}_h = \sum_h \tilde{\theta}_{ht_1}^* = \tilde{H}_{t_1} \quad (A11)$$

Equation (A12) expresses this in probability terms:

$$1 = \sum_h \frac{\tilde{\theta}_{ht_1}^*}{\tilde{H}_{t_1}} = \sum_h \tilde{\vartheta}_{ht_1}^* \quad (A12)$$

The basic mathematical problem is to derive the new household weight ( $\tilde{\theta}_{ht_1}^*$ ) or its probability distribution ( $\tilde{\vartheta}_{ht_1}^*$ ) from the initial weight ( $\theta_{ht_0}^*$ ) or distribution ( $\vartheta_{ht_0}^*$ ) and by imposing several constraints such as education, gender, or age-group population growth rates. Formally, the new probability distribution can be obtained by solving the optimization problem in equation (A13):

$$\min \Gamma(\vartheta, \tilde{\vartheta}) = \sum_h \tilde{\vartheta}_{ht_1}^* \ln\left(\frac{\tilde{\vartheta}_{ht_1}^*}{\vartheta_{ht_0}^*}\right) \quad (\text{A13})$$

subject to

$$\begin{aligned} E(x_{rt}) &= \sum_h \vartheta_{ht_0}^* k_{x_{rt}h} \quad r=1, \dots, R \\ \sum_h \vartheta_{ht_0}^* &= 1 \end{aligned}$$

The term  $k_{x_{rt}h}$  is the mean of the  $x_r$  characteristic within the household  $h$  ( $k_{x_{rt}h} = \frac{\sum_i x_{iht}}{hsize_h}$ ). In other words, this optimization procedure picks the best distribution  $\tilde{\vartheta}_{ht_1}^*$  that meets the moment constraints and the normalization restriction while, at the same time, requiring the least additional information and the least distance from the initial distribution ( $\vartheta_{t_0}^*$ ). The solution to this problem is given by equation (A14):

$$\begin{aligned} \hat{\vartheta}_{ht_1} &= \vartheta_{ht_0}^* \frac{\exp(\chi \hat{\lambda})}{\Omega(\hat{\lambda})} \rightarrow \sum_h \hat{\vartheta}_{ht_1} = \sum_h \vartheta_{ht_0}^* q_h = \tilde{H}_{t_1} \\ q_h &= \frac{\exp(X \hat{\lambda})}{\Omega(\hat{\lambda})} \end{aligned} \quad (\text{A14})$$

in which  $\hat{\lambda}$  is the Lagrange multiplier for the constraint  $E(X)$ , and  $\Omega$  is the normalizing factor to scale the new population to the target size.

#### Accounting for the reallocation of labor across economic sectors

The next step in the procedure is to account for the prospects of labor movements across economic sectors and over time as defined by baseline or counterfactual scenarios provided by the macro or CGE model. Initial labor supply,  $l_{pqt_0}$  and labor income,  $w_{pqt_0}$ , can be obtained from the household level data. Nevertheless, at  $t_1$  or after, workers will tend to move away from the shrinking sector into the expanding sector if this choice represents an increase in their expected earnings. In this regard, the GIDD implements a probabilistic way to select the migrants from the shrinking sector.

The probability of observing an individual  $i$  working in the expanding sector ( $p=p_1$ ) can be modelled as a probit equation (A15):

$$Pr(p = p_1) = P(X_i, Z_i) \quad (\text{A15})$$

where  $X_i$  and  $Z_i$  are vectors of individual and household characteristics, respectively, for individual  $i$ . Workers in their shrinking sectors,  $p_0$ , can be assigned a score based on their set of characteristics  $X$  and  $Z$ , and sorted accordingly. Workers in the shrinking sector,  $p_0$ , with the higher probability of being in the expanding sector,  $p_1$ , can migrate up to a point where the predicted share of labor is satisfied. That predicted share is a result from the macro model.

Equation (A16) describes the ratio  $\delta$  or the relative share of less qualified (i.e., less educated) labor,  $q_0$ , between the period  $t_0$  and  $t_1$  in a particular sector from the macro scenario. A value  $\delta > 1$  (a rise in the share) implies there is labor migration into the sector from the macro side.

$$\delta = \frac{\delta_{t_1}}{\delta_{t_0}} = \frac{\frac{l_{p_1 q_0 t_1}}{l_{p_1 q_0 t_1} + l_{p_0 q_0 t_1}}}{\frac{l_{p_1 q_0 t_0}}{l_{p_1 q_0 t_0} + l_{p_0 q_0 t_0}}} \quad (\text{A16})$$

This reallocation is fed into the microsimulation next. To do this, we note that the corresponding ratio of labor shares in the micro side is as follows:

$$\delta' = \frac{\delta'_{t_1}}{\delta'_{t_0}} = \frac{\frac{l'_{p_1 q_0 t_1}}{l'_{p_1 q_0 t_1} + l'_{p_1 q_1 t_1}}}{\frac{l'_{p_1 q_0 t_0}}{l'_{p_1 q_0 t_0} + l'_{p_1 q_1 t_0}}} \quad (\text{A17})$$

where the labor supply of workers is defined as:

$$\begin{aligned} l'_{pqt_0} &= \sum_h \sum_{i \in pqh} \theta_{ih} \\ l'_{pqt_1} &= \sum_h \sum_{i \in pqh} \tilde{\theta}_{ih} \end{aligned} \quad (\text{A18})$$

Whenever the condition  $1 \leq \delta' < \delta$  holds, it means the total number of workers in the expanding sector of the micro side has to be adjusted to match the macro scenario by allowing the new demographic structure to operate :

$$l^* = \delta'_{t_0} \frac{\delta_{t_1}}{\delta_{t_0}} (l'_{p_1 q_0 t_1} + l'_{p_1 q_1 t_1}) - l'_{p_1 q_0 t_1} \quad (\text{A19})$$

If  $l^* > 1$ , GIDD assumes that there is an extra amount of labor that can migrate and the likely candidates are determined by the probit function (A19). To avoid a larger sectoral reallocation than the CGE or macro results, the migration ceases when  $l^* \leq 1$ .

#### Accounting for changes in wages and skill premia by sector

The next step adjusts factor returns by skill and sector in accordance with the results of the CGE model. The GIDD imposes an entirely new vector of earnings on each worker, conditional on that worker being in sector  $p$  and qualification level  $q$ . In order to ensure consistency between the macro and micro outcomes, the GIDD adjusts the ratios between wage premia rather than wages themselves.

Beginning with a distribution of earnings from labor,  $w$ , by sector,  $p$ , and qualification,  $q$ , in the macro data, a set of wage ratios is formulated as follows:

$$g_{pqt_0} = \frac{w_{pqt_0}}{w_{p_0q_0t_0}} \quad (A20)$$

$w_{p_0q_0t_0}$  is the initial average earnings from labor for the less qualified workers in the shrinking sector. Similarly, a set of wage ratios for the macroeconomic counterfactual scenario can be defined as follows:

$$\hat{g}_{pqt_1} = \frac{w_{pqt_1}}{w_{p_0q_0t_0}} \quad (A23)$$

where  $w_{pqt_1}$  and  $w_{p_0q_0t_1}$  represent wage values from the CGE model in the macroeconomic counterfactual scenario. In general, the micro data will have a set of initial wage ratios  $g'_{pqt_0}$  that will differ from the macro data's  $g_{pqt_0}$ . Define:

$$g'_{pqt_0} = \frac{\sum_h (w'_{pqt_0} \cdot \vartheta_{ht_0}^*)}{\sum_h (w'_{p_0q_0t_0} \cdot \vartheta_{ht_0}^*)} \quad (A42)$$

where  $g'_{s,e}$  represent weighted wage ratio by skill qualification,  $q$ , and sector,  $p$ ,  $w'_{pqt_0}$  are labor earnings obtained from household survey data. Similarly

$$g'_{pqt_1} = \frac{\sum_h (w'_{pqt_1} \cdot \tilde{\vartheta}_{ht_1}^*)}{\sum_h (w'_{p_0q_0t_1} \cdot \varphi_{ht_1}^*)} \quad (A25)$$

The counterfactual wage ratios  $\hat{g}'_{s,e}$  for the microdata will then be calculated as:

$$\hat{g}'_{pqt_1} = g'_{pqt_0} \frac{\hat{g}_{pqt_1}}{g_{pqt_0}} \quad (A26)$$

This implies that even if initial and final wages differ between the macro and micro models, the change in wage ratios will be consistent across the two models. This diminish the possibility of a reversal in the wage gap between skill levels and ensures that the distributional changes are consistently mapped from the macro to the micro data.

### Accounting for economic growth

In order to adjust the micro data such that the weighted average percentage change in the per capita income/consumption across all households matches the change in real consumption per capita in the CGE model, a subsequent adjustment is carried out. Define  $Y$  as real per capita income calculated from the CGE model in the benchmark and  $\hat{Y}$  as its predicted value in the CGE model simulation. Define  $\gamma'_h = \sum_i y'_{i,h} / n_h$  as the per capita income of household  $h$ , where  $i$  identifies each household member, and  $n$  is equal to the size of household  $h$ .  $Y'$  is the weighted average value of real per capita income across all households, i.e.,

$$\sum_h v_h \gamma'_h = Y' \quad (A27)$$

where  $v_h$  is the weight of household  $h$  in aggregate income. Correspondingly

$$\sum_h \omega_h \lambda \hat{\gamma}'_h = \hat{Y}' \quad (\text{A28})$$

is the weighted average per capita income value in the policy simulation. Note that  $\sum_h v_h = 1$ ,  $\sum_h \omega_h = 1$  and  $\lambda$  is a scalar. Equations 26 and 27 allow for different household weights since the weights of the households will typically change over time. We constrain  $\hat{Y}'$  by equation 27:

$$\hat{Y}' = Y' \frac{\hat{Y}}{Y} \quad (\text{A29})$$

We implement this constraint in a distribution neutral way. That is, we adjust all household income in the counterfactual by a scalar  $\lambda$  such that per capita household income equals  $\lambda \hat{\gamma}'_h$ : as a result,  $\lambda$  can be defined by:

$$\lambda \sum_h \omega_h \hat{\gamma}'_h = Y' \frac{\hat{Y}}{Y} \quad (\text{A210})$$

This step in the microsimulation assures that national income grows accordingly to the results of the CGE model. Despite the fact that the GIDD ignores other forms of income, such as capital income (primarily because of missing data), this transformation guarantees consistency between the weighted average income in the household and CGE model assessment. For poor households, which is the main focus of the work, the assumption should be reasonably accurate since poor households have little capital income. There is more of a margin of error for wealthier households. Even so, for these households, it is skilled labor rather than unskilled labor that tends to be more important. For example, Bussolo, de Hoyos, Medvedev (2010) noted the tendency for skilled wage and returns to capital to be correlated.

#### **Accounting for the relative prices of food vs. non-food**

Given the importance of food consumption in the household budget for many developing countries, the GIDD accounts for the effects of food and non-food prices in household expenditures. As data improves in the household survey, more categories of consumption could be added. In the current framework, changes in agricultural and non-agricultural prices from the macro scenario are distributed across heterogeneous households using the following method.

Let us define the initial per capita monetary income of household  $h$ ,  $\gamma'_h$ , and the purchasing power of household  $h$ ,  $\gamma^r_h$ , as the ratio of its monetary income divided by a household-specific price index, capturing the household's consumption patterns in terms of food and non-food expenditure:

$$\gamma^r_h = \frac{\gamma'_h}{P_h} = \frac{\gamma'_h}{\alpha_h P_f + (1 - \alpha_h) P_{nf}} \quad (\text{A211})$$

where  $P_f$  and  $P_{nf}$  are food and non-food price indices and  $\alpha_h$  is the proportion of household's  $h$  budget spent on food. Equation 30 captures the dual effect of a price variation, i.e. higher monetary

income on the one hand, and the loss of purchasing power on the other. The  $\alpha_h$  parameter in the denominator of the right hand side is estimated from the following regression:

$$\alpha_h = \beta_0 + \beta_1 \ln(\gamma'_h) + e_h \quad (\text{A30})$$

where  $e_h$  is a vector of household-specific errors that are assumed to be distributed with  $E(e_h) = 0$  and  $V(e_h) = \sigma^2$ . Assuming that estimated parameters  $\hat{\beta}_0$  and  $\hat{\beta}_1$  remain constant, the new budget share spent on food for household  $h$ ,  $\alpha'_h$ , at the counterfactual per capita income,  $\lambda\hat{\gamma}'_h$ , can be obtained from:

$$\alpha'_h = \hat{\beta}_0 + \hat{\beta}_1 \ln(\lambda\hat{\gamma}'_h) + \hat{e}_h \quad (\text{A31})$$

The changes in real per capita incomes brought about by a change in relative prices of food versus non-food can be approximated by the following linear expression:

$$\hat{\gamma}_h^r = \frac{\lambda\hat{\gamma}'_h}{\alpha'_h P'_f + (1 - \alpha'_h)P'_{nf}} \quad (\text{A32})$$

where  $\hat{\gamma}_h^r$  is the real per capita income adjusted for changes in relative prices of food versus non-food.  $\hat{\gamma}_h^r$  is the counterfactual welfare aggregate for the analysis of poverty and shared prosperity.