

# Preface

The literature has identified several pitfalls in measuring the total distribution of income via household surveys, while substantially smaller problems are found for employee compensation. Similarly, a key challenge in measuring the labour income share is to estimate the labour income of the self-employed (half of the global workforce), in contrast to employee compensation. The labour income share and the labour income distribution play a key role in income inequality. Both are estimated, accounting for self-employment, based on observation-level employee compensation. This approach has often been considered appealing but not practical for a large panel of countries. Meanwhile, distribution measurement efforts have seldom focused on labour income, and yet it faces substantially less measurement challenges than total income and has a high policy relevance.

This analysis includes household surveys for 95 countries, spanning 2004-2017, mainly from the ILO Harmonized Microdata collection.<sup>1</sup> The labour income of the self-employed is estimated based on employees of similar characteristics. Afterwards, the labour income share, the capital share, and the labour income distribution are produced. A review of existing evidence concerning household surveys is undertaken, including a country case study comparing an establishment survey (with income data from the social security register) with a household one, and it is found that household surveys can provide reasonable estimates of the labour income distribution. The main results concerning the labour income share and distribution are: (i) the global labour income share is declining and countercyclical, similar patterns arise in the European Union and the United States; (ii) the effects of self-employment on the labour income share are highly heterogeneous – highlighting the limitations of widely used rules of thumb; (iii) labour income inequality decreases strongly with national income level, hence average cross-country differences are greatly exacerbated; (iv) within countries, relative increases of labour income at the upper end of the distribution are associated, on average, with relative losses for the rest.

**Keywords:** Labour Income Share, Inequality, Labour Income Distribution

The Labour Income Share and Distribution dataset was produced by the ILO Department of Statistics. This document was prepared by Roger Gomis (ILO Department of Statistics). The author thanks Steven Kapsos for his extensive guidance and support during this project and is grateful to Patrick Belser, David Bescond, William Cole, Rafael Diez de Medina, Rosina Gammarano Lame, David Kucera, Stefan Kühn, Sangheon Lee, Nicolas Maitre, Quentin Mathys, Marc Morgan, Yves Perardel, Marie-Claire Sodergren, Rosalia Vazquez-Alvarez, and Mabelin Villarreal Fuentes for their helpful discussions and comments.

---

<sup>1</sup> [https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms\\_651746.pdf](https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_651746.pdf)

# 1. INTRODUCTION

---

The measurement of the labour income share (LIS) is a topic of substantial interest in the field of economics. The stability of the labour income share, documented by Kaldor (1957), has recently captured attention following an apparent decline, as has the cyclical behaviour of the measure. Similarly, the apparent variation of the labour share between countries at different income levels has attracted intense scrutiny. Furthermore, the measure is an important macroeconomic variable, for instance it is used to estimate the New Keynesian Phillips curve, as in Galí and Gertler (1999). The labour income share has also captured attention outside the academic debate, particularly as an inequality measure. The measure is included as an indicator to measure progress towards the United Nations Sustainable Development Goals.

At the same time, since Gollin (2002) it has been clear that its measurement is not straightforward. The main problem relates to estimating the –counterfactual– labour income of the self-employed. Self-employment constitutes half of the global workforce and, given the negative relationship observed between self-employment and national income, these measurement problems have been highlighted mostly in developing countries. Nonetheless, the need to account for self-employment is widely acknowledged even in high income countries. This is the reason for the existence of two measures: the adjusted labour income share (adjusted for self-employment) and the unadjusted labour income share.

Two main strategies are frequently used to adjust the LIS: the mixed income approach and the self-employment approach. The mixed income approach is based on splitting the income of the self-employed, as measured by the national accounts mixed income item, between capital and labour. The second adjusts the labour income on the basis of the “compensation of employees” item of national accounts and on the self-employment rate in a given economy. Both approaches are widely used and present strengths and weaknesses. The first approach has as main limitations the measurement problems of mixed income and the split of the self-employment income between labour and capital. The main limitation of the self-employment approach is how to assign an amount of labour income to the self-employed, relative to the labour income of employees. Since the choice of the relative labour income of the self-employed can be informed with microdata, while the mixed income measurement problems are not straightforwardly fixed, the focus of this study is on the second method: the self-employment approach. Regardless of the data source used to do the adjustment, the literature has overwhelmingly favoured a rule of thumb approach to estimate the self-employed income. Taking advantage of a standardised microdata repository, the ILO Harmonized Microdata collection,<sup>2</sup> the methodology pioneered by Young (1995) is extended to substitute the rules of thumb by a (micro) data driven approach. This type of exercise has often been characterized in the literature as a best practice, but largely unattainable at the international level due to data constraints. With this new collection of harmonized microdata, the results for 95 countries can be directly estimated.

The estimation of the relative labour income of the self-employed is based on the observable characteristics of those workers and how they compare to employees. Relevant variables, such as economic sector, occupation, education and age, are used in a regression setting to study the determinants of labour income of employees. Based on the estimated relationship between labour income of employees and the explanatory variables, labour income is extrapolated to the self-employed. Additionally, a correction procedure is implemented to reduce the effect of selection bias in self-employment. Afterwards, the relative labour income can be directly computed by aggregating across the work force. The estimates of labour income for self-employment are highly heterogeneous across countries, over time, and within self-employment categories, thus the use of popular rules of thumb is

---

<sup>2</sup> The ILO Harmonized Microdata collection is complemented with the Luxembourg Income Study data for two countries.

prone to severe limitations. The results of the present study suggest that using rules of thumb underestimates adjusted labour income shares in developed countries, whereas in developing countries the opposite is true. Furthermore, given the prevalence of self-employment in poorer countries, the assumption of a rule of thumb for self-employment labour income completely drives the results of the estimation in these countries. Therefore, such estimates are not only affected by bias; they are almost determined ex-ante, greatly lowering their informational value. In contrast, the microdata approach derives results from labour income patterns observed in the data, such as how earnings change as a function of economic sector. The methodology outlined is not only limited to countries with available microdata, due to the imputation of missing data. The results show that there are exploitable patterns across national income, self-employment rate, and detailed status in self-employment, which allow to avoid rules of thumb even in countries with no microdata.

The microdata-adjusted global labour income share presents a downward trend in the period from 2004 to 2017, with a clear countercyclical behaviour during the financial crisis and in its aftermath. Furthermore, whereas level of per-capita income has a very mild relationship with the LIS across countries, not properly adjusting for self-employment can greatly overstate or understate the true association. The results also show that the microdata-based approach reduces the relative dispersion of the labour income share.

Adjusting the labour income share is a data intensive process, one that requires having an observed labour income or an imputed labour income for each observation unit (for each worker). These data are an interesting “by-product” that can be exploited for computing the distribution of labour income. Producing distributional data of total income, mainly in high income countries, has recently been a focus of interest in the academic literature. For instance Piketty, Saez and Zucman (2018) set a stepping stone to produce distributional national accounts. Similarly, in developing countries, distributional data on income or, preferably, consumption has been of great interest to produce poverty estimates, for instance in Chen and Ravallion (2010). Both lines of work, the distributional national accounts and the international poverty estimation, have seen a massive data production effort, the results of which can be found, among others, in the World Inequality Database (<https://wid.world>) and in the POVCALNET repository of the World Bank (<http://iresearch.worldbank.org/PovcalNet>).

The role of income distribution in the world of work, outside the labour income share, has not drawn similar levels of interest, and has not seen much innovation vis-à-vis data collection and data production activities. Nonetheless, many of the debates around inequality and poverty, both in the literature and the society as a whole, seem to be closely related to the world of work. For instance, consider the role of the following factors, which are active areas of both academic and social debate, in inequality and poverty: job polarisation, automation, minimum wages, gender wage gaps, aging populations, new forms of work and the gig economy, the decline of the middle class, international trade and globalisation, or the decline of unions. These topics can certainly be considered as potential drivers of labour income and its distribution. Therefore, this study and the related data production exercise aim to palliate the data scarcity specifically related to the distribution of labour income. All the estimates, which include the labour income distribution by percentile at the country level, have been made publicly available in ILOSTAT, the ILO’s repository of international labour statistics.

From the methodological point of view, the current study sheds light into harmonized labour survey data, which constitute a rich and internationally comparable data source to study the world of work. Furthermore, the focus on labour income adds a new perspective into an important topic in the inequality literature: the suitability of using survey data to compute inequality metrics. Piketty, Saez and Zucman (2018) argue that administrative data present an advantage over household survey data to compute the income distribution due to, mainly, under coverage of high incomes in surveys. This topic is revisited focusing on labour income, as opposed to overall income. This new perspective points to the convenience of a systematic cross check to evaluate whether household surveys are accurate enough to

measure wage distributions, even if they are not accurate enough to measure overall income distribution. The natural benchmark in such a process would be establishment surveys and administrative/tax registers. Due to data and metadata restrictions, the present study does not undertake such an exercise. Nonetheless, a country case study for Belgium, comparing the distribution of employee compensation from a household survey with the one from an establishment survey (that randomly draws earnings data from the social security register) shows very encouraging results. The employee compensation distribution produced by the household survey is very close to its benchmark. In a similar manner, existing evidence, both from the literature and published by an official statistical agency (Eurostat), points to a high degree of consistency between labour survey data and national accounts in terms of coverage rates (the ratio of aggregate labour income by type of source). Moreover, the gaps in coverage rates (the mismatch between sources) do not appear to systematically bias the labour income distribution in terms of over- or underestimating inequality, in contrast to overall income. The results of the case study benchmarking a household survey to an establishment survey, the reported high coverage rates, and the finding that coverage gaps are not systematically associated with inequality over- or underestimation provide a highly encouraging, yet tentative, methodological justification for the use of household surveys to estimate labour income distributions.

From the perspective of the results, it is interesting to highlight two facts concerning the labour income distribution. First, the degree of labour income inequality is strongly and negatively correlated with national income; this exacerbates average income differences between countries, resulting in a very unequal distribution of global labour income. In 2017 the top decile earned around half of the overall global labour income, whereas the poorest half of the workforce earned less than seven per cent. Second, within countries, increases in the shares of labour income for the highest earners are associated with decreases for much of the other workers, whereas increases in the median share of income tend to be positively associated with increases in much of the distribution, particularly towards the lower end, and decreases for the top. Interestingly, in some countries like the United States and Germany, this pattern is not found. Instead the middle class experiences losses while gains occur at the upper end, whereas the poorest percentiles do not experience losses and even make gains in relative terms.

The article is organized into four sections. Section II describes in detail the methodology and its relationship to the literature. Section III presents some of the results, including the global labour income share and distribution and their evolution since 2004. Section IV concludes.

## 2. METHODOLOGY

---

This section describes the estimation of the labour income share and its distribution at the country, regional and global levels. In the appendix, further details concerning the sources of the data can be found. This study combines national accounts data with labour survey data<sup>3</sup> to compute the labour income share and its distribution. In particular, starting with variables from the System of National Accounts (SNA), adjustments and distributions computed from labour survey data are applied to obtain the estimates.

### 1. THE UNADJUSTED LABOUR INCOME SHARE: THE STARTING POINT OF THE ADJUSTED LABOUR INCOME SHARE

The first step to estimate the labour income share is to compute the unadjusted labour income share from national accounts data. This measure is defined in equation (1):

$$\text{Unadjusted LIS} = \frac{\text{Compensation of Employees}}{\text{GDP}} \quad (1)$$

The denominator of the expression corresponds to the item B1g of the SNA 2008, the gross domestic product (GDP). This magnitude is commonly used in the LIS empirical literature as the adequate scaling factor to benchmark labour income, see for instance Gollin (2002), IMF (2017), Cho, Hwang and Schreyer (2017). An important exception is Bridgman (2014), who subtracts from GDP the depreciation of fixed capital. In the literature concerning income distribution adjusting for fixed capital depreciation is also common, for instance see Piketty, Saez and Zucman (2018). This is appealing since depreciation does not accrue to capital owners. Nonetheless, in this study GDP is used for the following practical reasons. First, as described in the SNA 2008 handbook, UNSD et al. (2009), “Consumption of fixed capital is one of the most difficult items in the accounts to define conceptually and to estimate in practice”. This difficulty does not only increase the uncertainty of the estimate, in many countries it directly causes the item to not be computed, or at least published. Second, the focus of this study is labour income and its distribution. The fact that the capital share of GDP includes capital income plus depreciation does not inhibit analysis of the remaining labour component. Third, given the objective of obtaining global and regional estimates of the labour income share and its distribution, a price level adjustment of country GDP is necessary. Conveniently, GDP adjusted by PPP is widely available. Fourth, this study introduces several new elements in estimating the labour income share compared to the previous literature. In this setting it is convenient to use GDP as it is a measure that has been

---

<sup>3</sup>The term labour survey data is used (as a term analogous to national accounts data) to refer to survey data that broadly follow the International Conference of Labour Statisticians (ICLS) standards on labour statistics. The ICLS participants include experts from governments, mostly appointed from ministries responsible for labour and national statistical offices, as well as from employers’ and workers’ organizations. The International Conference of Labour Statisticians is invited to make recommendations on selected topics of labour statistics in the form of resolutions and guidelines, which are then approved by the Governing Body of the International Labour Organization before becoming part of the set of international standards on labour statistics. These standards usually relate to concepts, definitions, classifications and other methodological procedures which are agreed as representing ‘best practice’ in the respective areas, and which, when used by national producers, will increase the likelihood of having internationally comparable labour statistics as well as comparability across time within a country.

commonly used before.<sup>4</sup> Nonetheless, it should be highlighted that throughout this study a reference to capital income or the capital share includes the item of fixed capital depreciation.

The numerator of the expression corresponds to item D1 of the System of National Accounts 2008, UNSD et al. (2009). Compensation of employees<sup>5</sup> (D1) is defined as “as the total remuneration, in cash or in kind, payable by an enterprise to an employee in return for work done by the latter during the accounting period.” This definition includes both the wages received by the employee and the employer’s social contributions made on behalf of the employee. Using the compensation of employees as a lower bound of total labour income is a widely used practice (an early example would be Johnson (1954)), documented in detail in the seminal work of Gollin (2002). The measure can be taken as a lower bound to labour income since it only includes the earnings of employees. Thus the unadjusted labour income share ignores the labour income of the self-employed.

The definition of self-employed can be found in the revision of the International Classification of Status in Employment, ILO (1993), and includes – among others – employers, own-account workers and contributing family workers.<sup>6</sup> Employers are self-employed individuals who engage at least one employee in a regular basis. In contrast, own-account workers do not engage employees on a regular basis. Finally, contributing family workers work in an establishment, operated by a relative, with a limited degree of involvement in its operation. Both employees and the self-employed, as well as their subcategories, can be readily obtained from labour survey data. From these definitions, it is clear that such individuals provide labour to the business that they work in and hence generate (in a direct or indirect manner) labour income. Employers work on their business and the income that they extract from it will in part include remuneration for labour as well as capital. The same circumstance applies to own-account workers. The share of the business income that accrues to labour will depend on the circumstances of each individual producer, but it is reasonable to consider that the share of labour income can be substantial. Finally contributing family workers, whilst not receiving any direct pay, generate income for the family business which in turn results in household income. At least some of this income, and arguably a substantial amount, is the product of labour input and therefore should be assigned to labour income. Hence, across any major sub-category of self-employment, it is reasonable to assume the production of a substantial amount of labour income.

## **2. THE MISSING PIECE OF THE UNADJUSTED LABOUR INCOME SHARE: THE LABOUR INCOME OF THE SELF-EMPLOYED**

Given that the SNA item compensation of employees generally disregards the self-employed, the unadjusted labour income share will be a downwards-biased estimate of the labour income share. The reason is that GDP (and other related measures) includes the income of the self-employed. Part of the income generated by self-employed individuals is bound to be labour income, as discussed above, while

---

<sup>4</sup> Piketty, Saez and Zucman (2018) use national income instead of domestic income as a reference to produce distributional estimates. The reason is clear, foreign income is sizeable for the capital income of top earners. However, the national measures are less widely available than domestic ones (including the PPP adjusted measures).

<sup>5</sup> The definition of employees in the SNA is broadly consistent with the Revision of the International Classification of Status in Employment (ICSE), ILO (1993), approved by the 15th International Conference of Labour Statisticians (ICLS). This is relevant since the status in employment used in surveys follows this framework. Nonetheless, even if it is broadly consistent, some discrepancies might arise. In the appendix a discussion concerning this issue can be found. It is worth highlighting that a new ICSE was approved during the 20<sup>th</sup> ICLS, but it has yet to be implemented.

<sup>6</sup>There is a fourth category, members of producers’ co-operatives, which is scarcely used in practice.

the rest would accrue to capital. As long as there is a positive share of the self-employed income that accrues to labour the unadjusted labour income share will underestimate the true labour income share.

This circumstance is very relevant at the empirical level and the necessity to adjust this biased measure is widely recognized in the literature,<sup>7</sup> see for instance Krueger (1998) and Gollin (2002). This necessity has been highlighted particularly in low and middle income countries. These two facts are not at all surprising, according to ILO (2018a) the global average prevalence of self-employment during the last 20 years is close to one-half. In low income countries, around four-fifths of the workforce are self-employed. In high income countries the rate is close to one in ten. Therefore, the computation of the labour income share and its distribution is made more complex, requiring data on the labour income of the self-employed. Theoretically, the income of the self-employed will contain both labour and capital income. Empirically, however, there is not a completely satisfactory way to estimate it, which has led to multiple proxy approaches in the literature.

Many studies take the strategy to estimate the counterfactual labour income of the self-employed based on the mixed income of the System of National Accounts. According to UNSD et al. (2009) mixed income is the magnitude accrued from production of unincorporated enterprises owned by households. As such, it implicitly contains unpaid labour inputs provided by the owner or their family members. In fact, the SNA framework recognizes that the remuneration of labour may dominate the overall value of mixed income.<sup>8</sup> Several corrections to the unadjusted labour income share on the basis of mixed income have been proposed. All the proposals have in common that they treat compensation of employees as a lower bound for total labour income and mixed income as an upper bound for the labour income of the self-employed (the second one is a consequence of assuming that mixed income reflects all of the income of the self-employed). Whereas the first assumption is plausible, the second one – given the measurement problems concerning mixed income described below – is at the very least worth re-examining. The nomenclature below follows Gollin (2002) for its synthetic value. Nonetheless, it should be noted that, while widely used, the terminology varies substantially in the literature. Furthermore, the approaches of Gollin are parametrised to make easier the comparison of different proposals in the literature. Most of the corrections based on mixed income can be classified as broadly following Gollin’s (2002) first approach. The first correction, labelled G1, is:

$$G1 LIS = \frac{\text{Compensation of Employees} + \gamma \cdot \text{Mixed Income}}{GDP}; \gamma = 1 \quad (2)$$

The inclusion of  $\gamma = 1$  is aimed at facilitating the comparison of other related proposals made in the literature. This equation simply assumes that all income of unincorporated businesses owned by households is labour income. Based on the notion that mixed income is the upper bound of the self-employed labour income this measure is often described as an upwardly biased measure of the labour income share. In the literature several modifications of  $\gamma$  (which simply reflects the share of mixed income that is considered labour income) are proposed. In the same work Gollin (2002) proposes a second approach:

$$G2 LIS = \frac{\text{Compensation of Employees}}{GDP - \text{Mixed Income}} \quad (3)$$

Notice that Equation (2) and (3) can be made equivalent if  $\gamma$  is assumed to be:

$$\gamma = \frac{\text{Compensation of Employees}}{GDP - \text{Mixed Income}} \quad (4)$$

<sup>7</sup> In fact, this is the reason that the ratio between compensation of employees and GDP is known as the “unadjusted” measure.

<sup>8</sup> The term “mixed” indicates that the item is composed by both labour and capital income.

This simply states that the share of labour income in the mixed income item is the same as in the employee sector. Therefore, both G1 and G2 can be considered in the same class of modifications. The G2 approach will yield an unbiased measure of the labour income share as long as two assumptions hold: mixed income reflects all the income of the self-employed and the labour income share is the same in the employee sector and in the self-employed sector. Whereas the validity of the first assumption is dependent on measurement practices (which are discussed below) but appealing in theory, the second one is problematic, given that it partly assumes the result of the adjustment.<sup>9</sup> Can all the adjusted mixed income measures be mapped - with the appropriate gamma - to the G1 equation? The answer is broadly yes, albeit often the treatment of the taxes included in GDP varies from study to study. To mention a few: the methodology described in Appleton (2011) and ONS (2018) used by the UK's Office for National Statistics would be in line with G2, as would be Karabarbounis and Neiman (2014) or Piketty (2014). Johnson (1954) used a variant of G1 with  $\gamma = 2/3$ . In the Penn World Tables v8 several adjustments based on mixed income are used including G2 when mixed income is available, as well as a variant of G1 in which the labour income of the self-employed is approximated as the value added in agriculture, see: Feenstra, Inklaar and Timmer (2015).

The measurement of mixed income and the assumption that it adequately represents total income of the self-employed (not only the labour income) are subject to a few problems. First, mixed income can be relatively hard to measure, particularly in low and middle income countries, resulting in an absolute lack of data for many countries. Second, even within the SNA framework there is a substantial heterogeneity in practices in country national accounts, particularly in the estimation of mixed income. This is not limited to the developing world. In just four OECD countries (France, Germany, Italy, and the United States), three of which follow the European System of National Accounts, Pionnier and Guidetti (2015) find substantially different national accounts practices with respect to mixed income. In particular, in Germany and Italy, a large part of the self-employed are classified as in the corporate sector and not in the "mixed income sector". In contrast, in France and the United States this is not the case, producing a very large disparity in the actual economic coverage of the item mixed income. In the case of Germany and Italy, therefore, mixed income can hardly be treated as an upper bound of self-employment income, given that a large share is accounted for in the corporate sector. Third, mixed income is notoriously prone to be underestimated due to underreporting compared to other items of the national accounts. For instance, an assessment provided by an Australian Bureau of Statistics official, Johnson (2003), provides estimates of the upper bound of underground activity missing from GDP. As upper bounds the author suggests: 4.8 per cent for overall GDP, 2.0 per cent for compensation of employees, 2.6 per cent of gross operating surplus, and 37.4 per cent for mixed income. This suggests a risk of underestimation 20 times larger in mixed income than in compensation of employees, and one order of magnitude above the one of GDP. More recently, Piketty, Saez and Zucman (2018) highlight that the allowances (from national accounts) for misreporting in the United States in 2013 were US\$80 billion and US\$538 billion for wages and unincorporated business profits respectively, which would represent a share of the total of 0.9 per cent and 39 per cent. It is worth highlighting that this points to a degree of misreporting in mixed income 40 times larger than in compensation of employees in a country with a well-developed tax system. One can infer that the situation in countries with less developed tax systems – and characterized by a small employee sector and a large self-employment sector – the disparities could be much larger. In conclusion, mixed income is a measure subject to a substantially higher error than other items of national accounts, with severe problems of data availability and international comparability, which can exclude large shares of reported self-employment income, and with a substantial risk of being underestimated due to under-reporting.

---

<sup>9</sup> It is equivalent to ignore the self-employed sector and focus simply on the employee sector. Nonetheless it is deemed preferable to the unadjusted measure given that the latter has a clear downward bias.



Given the limitations in both measuring mixed income as well as allocating a share of it to labour income, Gollin (2002) proposes another correction. The third correction, labelled G3, does not use mixed income data. Instead, it relies directly on the share of self-employment in an economy measured by labour survey data. The measure reads:

$$G3 LIS = \frac{\text{Compensation of Employees}}{GDP} \cdot \frac{\text{Share of Employees} + \gamma \cdot \text{Share Self employed}}{\text{Share of Employees}}; \gamma = 1 \quad (5)$$

As in the case of G1 the inclusion of  $\gamma = 1$  is done to be able to group together several proposals made in the literature. The G3 measure will be an unbiased indicator of the labour income share as long as, on average, the self-employed earn the same labour compensation as employees. More generally, for other variants of the adjustment, G3 will be unbiased as long as on average self-employed workers earn  $\gamma$  units of labour income per each unit earned by the employees. The adjustment with  $\gamma = 1$  is widely used for high income economies, for instance the European Commission's AMECO database provides adjusted labour income share for European Union members and other high income countries on the basis of this adjustment. The International Monetary Fund, in chapter three of the World Economic Outlook April 2017, also uses this adjustment.

The assumption of equal compensation has shown its limitations, particularly outside of high income countries. Worldwide, employees are much less likely than the self-employed to work in the informal sector, which in turn is associated with the highest share of working poor ILO (2018b). Furthermore, the two largest categories of self-employment, own-account and contributing family work, are often associated with vulnerability and lower working and living conditions Gammarano (2018). For instance, in a selected sample of 39 developing countries, Kapsos and Bourmpoula (2013) find that own-account workers and contributing family workers are dramatically over-represented in working poverty and extreme working poverty (corresponding to the international poverty lines), whereas employees are over-represented in economic classes above the poverty lines. Therefore, assuming the same earnings as employees will plausibly generate a strong upward bias in the labour income share in low and middle income countries. Furthermore, in these countries the share of employees can be very low, which, coupled with the same earnings assumption, can create extremely high adjusted labour income shares.<sup>10</sup> Similarly, as Treeck (2017) highlights, in the context of high income economies one can expect that self-employment is more associated with entrepreneurial activities with potentially higher earnings, and less a reflection of lower job opportunities. To capture this duality the author proposes to assume that the self-employed<sup>11</sup> earn, in labour income, a third less than employees in low and middle income countries, or that  $\gamma = 0.67$ . A similar strategy is followed by Cho, Hwang and Schreyer (2017), who assume that the self-employed earn half on average than employees ( $\gamma=0.5$ ). However, it is worth stressing that in the latter case this is applied to high income countries.<sup>12</sup>

The G3 approach, including its variants as Treeck (2017), allows to avoid the problems associated with measuring mixed income as well as equating it to self-employment income, outlined above. At the same time, the G3 approach raises the question of what is the appropriate ratio of average labour earnings between the self-employed and employees. Henceforth this magnitude is referred to interchangeably as *relative wage, relative imputed wage,  $\gamma$ , self-employment penalty or premium*.<sup>13</sup> The need to pin down this relative wage has resulted in some criticism concerning the whole G3 approach. Indeed, it is a complex question to assign a wage to the self-employed. Rule of thumb assumptions are admittedly

<sup>10</sup> For instance, in India the direct G3 measure greatly exceeds 100 per cent - which is not possible.

<sup>11</sup> The study in question uses agricultural employment as a proxy for self-employment.

<sup>12</sup> In the following section strong evidence is shown against the downward correction of self-employment labour earnings in most high income countries.

<sup>13</sup> This admittedly diverse nomenclature is adopted for ease of exposition in a given context.

blunt and with a scarce empirical basis. Nonetheless, it is worth highlighting that this is exactly the case with the G1 (or G2) variants as well. These latter approaches require rules of thumb, not on the relative wage of the self-employed, but on the share of labour income that the self-employed earn as proportion of total income. In principle, the relative wage of the self-employed is no harder to pin down than the labour income share of the self-employed. Furthermore, the G1 and G2 approaches apply their rules-of-thumb to the mixed income item of SNA, whereas the G3 approach relies on compensation of employees. As outlined above, the latter is more frequently available and more likely to be correctly measured and to match the concept of income that is targeted. Thus when considering an adjustment based on mixed income (following the G1 or G2 approach) or on compensation of employees (following the G3 approach), this study favours the latter.

The estimates of the labour income share based on the G3 adjustment and using rules-of-thumb are nonetheless quite limited. Several problems can readily be identified. First, the relative wage of the self-employed has a huge impact on results and its value is directly assigned by the researcher. Whether it is chosen to be one, two-thirds, or one-half, the prior assumption largely drives the results, greatly limiting the research value of the estimates. Second, the one-size-fits-all approach completely disregards the complexity of self-employment between countries. Even the proposals to assume one coefficient for developing countries and one for developed countries assume that economic and social development is a binary status with a sharp threshold, which is not present in many widely used economic or social indicators. Third, within self-employment, the relationship with vulnerable employment and hence with work conditions such as labour income is complex. Indeed, the main sub-components of self-employment, employers, own-account workers and contributing family workers face very different circumstances, which can be taken into account to provide a more precise estimate of the relative wage of the self-employed. And finally, the rule of thumb approach completely disregards the evolution of the relative wage across time. In low and middle income countries, where a secular increase in the share of employees and a downward trend of agricultural employment are taking place, the assumption of stability seems far-fetched. This is also the case in high income countries, where the financial crisis and its aftermath, as well as aging population and the rise of new forms of work, ILO (2019), De Stefano (2016), certainly have the potential to have a particular effect on self-employment.

In conclusion, the present discussion strongly argues in favour of estimating the labour income share on the basis of compensation of employees and the share of self-employment, and not on mixed income. Nonetheless, the need to go beyond rules of thumb approaches – in either case <sup>14</sup> – is even more pressing. The next sub-section focuses on providing a data-driven alternative.

### **3. ESTIMATING THE LABOUR INCOME OF THE SELF-EMPLOYED BEYOND RULES OF THUMB: A (MICRO) DATA-DRIVEN APPROACH**

In this section, the theoretical and practical aspects of estimating the labour income of the self-employed are described. Consistent with the discussion above, the focus is on estimating the ratio of the average labour income of the self-employed over the average labour income of the employee, or the relative wage.

The objective of the methodology is to avoid relying on rule-of-thumb approaches to incorporate self-employment labour income. Instead, an empirical approach that exploits existing data sources is used to produce the estimates. Labour survey data provides a unique opportunity to empirically inform the estimate of the relative wage of the self-employed. This study uses a large number of labour force and household surveys that broadly follow standards set by the International Conference of Labour

---

<sup>14</sup> The methodology from this study could be extended to adjust the labour share on the basis of mixed income instead of self-employment.

Statisticians (ICLS) on labour statistics. Specifically, from the ILO Harmonized Microdata collection which is now described. These surveys are used to produce – amongst others – labour market indicators such as labour force, employment, or unemployment published regularly by national statistical offices and international organizations. One advantage of this data source is that it follows a coherent set of definitions and procedures outlined by the ICLS standards which render the data internationally comparable.<sup>15</sup> The second advantage is that microdata, data in which the disaggregation is at the individual (worker) level, is available for labour survey data.

## 1. The ILO Harmonized Microdata collection and other microdata sources

The main source of microdata for this study is the ILO Harmonized Microdata collection. This repository is the largest internationally-comparable collection of labour-related survey datasets. As of May 2019, the repository was comprised of 11,819 datasets,<sup>16</sup> including 8,325 labour force survey datasets across 151 countries. These are mainly comprised of the official surveys conducted by national statistical offices. Each survey includes anonymized information on thousands of individuals in the surveyed households and on the households themselves. The advantage of this repository is that the data treatment and definitions of the variables are standardized, providing comparability both across countries and over time, as well as economies of scale when analysing the data.

The selected datasets range from 2004-2017, given that before 2004 the availability of the variables of interest – particularly income – is much lower. Furthermore, only surveys with labour income data for employees as well as hours worked data are used. The resulting selection –after manual data cleaning– is composed of 711 datasets, covering 93 countries, with a total 63 million individual observations. In addition, for the United States and Germany, datasets from the Luxembourg Income Study Database are used. The variables selected in each survey (if available) to carry out the analysis are: labour related earnings of employees, hours worked, economic activity and occupation of the worker, rural or urban residence, and key demographic variables: age, gender and education level. A description of the methodology used to process the microdata as well as the available variables can be found in ILO (2018c). Because of the key methodological role of the labour earnings of employees, the concept is discussed in depth. The relevant internationally comparable definition of labour related earnings of employees<sup>17</sup> is:

All receipts and benefits in cash, kind or services, which accrue, over a given reference period, to persons in paid employment, for themselves or in respect of their family members, as a result of their involvement in paid employment jobs. Such receipts and benefits may be paid by the employer, social security and insurance schemes or the State, in so far as they are derived by virtue of the employment status.

“Resolution concerning the measurement of employment related income” adopted by the 16<sup>th</sup> International Conference of Labour Statisticians (October 1998).

The resolution thus aims to capture the totality of the income generated by being employed, including social security and related schemes whether public or private.<sup>18</sup> The reference period recommended by the resolution is long, such as a full year, but it can be substantially lower. It must be highlighted that not every survey in the sample strictly follows the definition, nonetheless the existence of the standard

---

<sup>15</sup> It has to be highlighted that there can be, and there are, limitations to the international comparability arising from not adhering strictly to the ICLS standards as well as the flexibility of these standards to adapt to national circumstances.

<sup>16</sup> At the monthly, quarterly and yearly frequencies.

<sup>17</sup> For the sake of simplicity throughout this study, labour related income data is referred to as: wages, earnings, compensation or labour earnings.

<sup>18</sup> These contributions are included as well in the national accounts item of “Compensation of employees”.

and the frequent adherence to it produces a reasonably homogenous set of results, apt for international comparisons. The degree of heterogeneity of the data has to be compared to the one of similar existing data collections. In employment income, one of the harshest deviations that one can envisage at the measurement level is that a country collects gross compensation (following the ICLS standards) whereas another choses to collect net wages. This certainly would alter the distribution, and indeed there are such cases in the current dataset. Contrast this however, to the extremely popular distributional data repository POVCALNET from the World Bank. The target of POVCALNET data is to measure poverty, and hence when possible it focuses on household consumption distribution measures. If no such data is available, the distributional measures are based on household income. Both at the conceptual and empirical level it is plausible that the heterogeneity caused by the use of gross or net labour income is not greater than the one caused by the use of household consumption or income data.

## 2. Estimating the relative wage for the self-employed using microdata

In the empirical literature concerning the labour income share, the concept of using microdata to estimate the labour income of the self-employed is not new. Young (1995) produces a relative wage using the wages of employees and imputing them to the self-employed according to the economic activity, gender, age and education. This approach is often cited as a best-practice. For instance, in Gollin (2002) it is said:

Perhaps the best approach is that of Young (1995), who imputed wages to the self-employed in Hong Kong, Singapore, and South Korea on the basis of their sector, sex, age, and education. [...]There are problems with this procedure, of course: it is difficult to control for unobservable differences in entrepreneurial ability, and it is difficult to know how to treat returns to entrepreneurial ability. Moreover, it requires detailed micro data, which makes it difficult to do for a large sample of countries. Nonetheless, Young's approach gives a plausible way of estimating labor shares in economies with large numbers of self-employed people.

More recently Cho, Hwang and Schreyer (2017) comments that “[t]he theoretically most compelling approach is a procedure based on matching micro-data records at national level.” Despite the positive opinion on this type of procedure it is seldom carried out due to its data requirements. This study builds on the approach of Young (1995)<sup>19</sup> and the ILO Harmonized Microdata collection to apply this approach to an international panel data set.

The idea underlying the procedure is that since we can observe the wages of employees we can impute a counter-factual wage to the self-employed based on common observable characteristics. The first step is to run an OLS regression to estimate the determinants of the (log) wages of employees:

$$\ln(\text{hourly\_wage}_i) = \alpha + x_i' \cdot \beta + \varepsilon_i \quad (6)$$

The control variables,  $x_i'$ , are indicated in Table 3 in the appendix. Based on the estimated regression coefficients, an imputed wage is computed for all workers, regardless of status in employment. The imputed wage (IW) is simply:<sup>20</sup>

$$IW_i = \exp(\hat{\alpha} + x_i' \cdot \hat{\beta}) * \text{hours worked} \quad (7)$$

The set of imputed wages can be used to produce a first step estimate of the log of average imputed wage, relative to employees, for each group of the self-employed, or without loss of generality for own-

<sup>19</sup> Nonetheless, several methodological elements differentiate the Young study from the present one.

<sup>20</sup> To increase the use of information available, a variant of this imputation with the monthly wage, instead of the hourly wage, is used for workers with no hours of work data.

account workers. The first step estimate reads (where OAW is the abbreviation of own-account workers and EES the one of employees):

$$\delta_{OAW} = \frac{\ln(\text{mean}(IW_i | \text{status} == OAW))}{\ln(\text{mean}(IW_i | \text{status} == EES))} \quad (8)$$

which is simply the ratio in logs between the average imputed wage of own-account workers and the average imputed wage of employees (the measures for other groups of the self-employed such as employers and contributing family workers can be produced in an analogous manner). If there was no selection bias, we could obtain the relative wage of the own-account workers using the formula:

$$IW(OAW) = (\text{mean}(IW_i | \text{status} == EES))^{\delta_{OAW}-1} \quad (9)$$

However, this measure is likely to be affected by a selection bias of the self-employed. For instance, by non-measured ability. Indeed, it is plausible that in certain countries self-employment is associated with higher than average ability. For instance, in the context of developed countries, in which a very large share of the work-force consists of employees, it can be that many of those with above average ability choose self-employment as a way to increase earnings. Similarly, it is also possible that in developing countries, where jobs as an employee are scarce and better paid than self-employment, employees present a higher than average non-measured ability. Many other forms of selection bias can arise.<sup>21</sup>

To account for selection bias in self-employment one cannot use the popular two-step procedure based on Heckman (1979). The reason is straightforward; we do not have any observation of a self-employed wage simply because it is a counterfactual. Therefore, the correction based on the Heckman procedure cannot be produced. Nonetheless, the likelihood of a selection bias makes it advisable to at least mitigate the problem. With this objective, a second step estimate based on the following equation is used:

$$\theta_{OAW} = \delta_{OAW} + (\delta_{OAW} - 1) \cdot \left( \frac{1 - R^2}{R^2} \right) \quad (10)$$

Where  $R^2$  indicates the share of variance explained by the regressors in the wage regression. The imputed wages of the own-account workers can be computed as  $IW(OAW) = IW(EES)^{\theta_{OAW}-1}$ . The correction proposed to produce the second step estimate is based, first of all, on reframing the selection bias of the self-employed as an omitted variable bias. In this case, the omitted variables will not only cause a bias in the estimates of the self-employed wages, but will also manifest in lower r-squared in the regression performed on the wages of employees. As more of the omitted variables are added, the bias of the imputed wage of the self-employed will be reduced and the r-squared will be increased. Under this condition, the relative wage is extrapolated until no omitted variable could possibly exist in the employee regression, when the r-squared reaches 100 per cent.<sup>22</sup> The particular functional form, a linear extrapolation of the ratio in logs, has been adopted based on the following exercise. Aside from running the regression analysis on the wages of employees with the full set of explanatory variables, the results of a nested model with fewer variables can be collected. Hence, based on this restricted regression, a different imputed wage for the self-employed can also be produced, and a restricted ratio in logs can be computed, denoted  $\delta_{OAW}^R$ . Finally, a completely restricted regression (with no explanatory

<sup>21</sup> For instance, it is likely that employment status is endogenous to shocks to both earnings and productivity.

<sup>22</sup> In the presence of random noise, the point where there are no omitted variables left out of the regression can be in fact much lower than the 100 per cent threshold – unfortunately the share of variance due to noise is unknown. Nonetheless, the procedure is not drastically affected by the assumption of lower thresholds such as 85 per cent.

variables) can be notionally considered. In this case, the relative imputed wage would be 1 and the r-squared 0, this serves as an initial point to extrapolate the values.

The functional form is based on the observation that if we linearly extrapolate the restricted first step estimate,  $\delta_{OAW}^R$ , based on the r-squared of the regression,<sup>23</sup> to approximate the value of the full regression estimate,  $\delta_{OAW}$ , the imputed value is not far away from the actual value.<sup>24</sup> The extrapolation to approximate the value reads:

$$\delta_{OAW} \approx \delta_{OAW}^R + (\delta_{OAW}^R - 1) \cdot \left( \frac{R^2 full - R^2 restricted}{R^2 restricted} \right) \quad (11)$$

The proposed correction follows the same logic. In this case however, it is assumed that a regression with a 100 per cent r-squared will not present omitted variable bias. Hence, the value of the full regression r-squared is substituted by 1. Similarly, the model with all the included variables is considered as having some variables missing (the omitted variables), so in the place of the r-squared from the restricted regression the one from the full regression is used. This gives an expression identical to the one outlined in in (10):

$$true\ value \approx \theta_{OAW} = \delta_{OAW} + (\delta_{OAW} - 1) \cdot \left( \frac{1 - R^2 full}{R^2 full} \right) \quad (12)$$

The correction proposed then relies on the crucial assumption that the effect of truly omitted variables will affect the imputed wage of the self-employed to a similar extent as the pseudo omitted variables.<sup>25</sup> Given that the wage of the self-employed is unobservable, the results cannot be validated in an actual pseudo out-of-sample exercise. Unreported simulations suggest that the correction captures the direction of the bias well, but the accuracy of the approximation varies substantially depending on the particular specification of the model and the selection bias. The advantage of the proposed method is that it attempts to correct the selection bias,<sup>26</sup> using the information from the microdata, that it is likely to be present in a manner that –in simulations– seems to capture the direction of the bias. The disadvantage is that results depend highly on the specific form of the data generating processes assumed, which in the context of this application we cannot observe directly.<sup>27</sup> Figure 1 illustrates an example of the procedure, based on an idealized example.

---

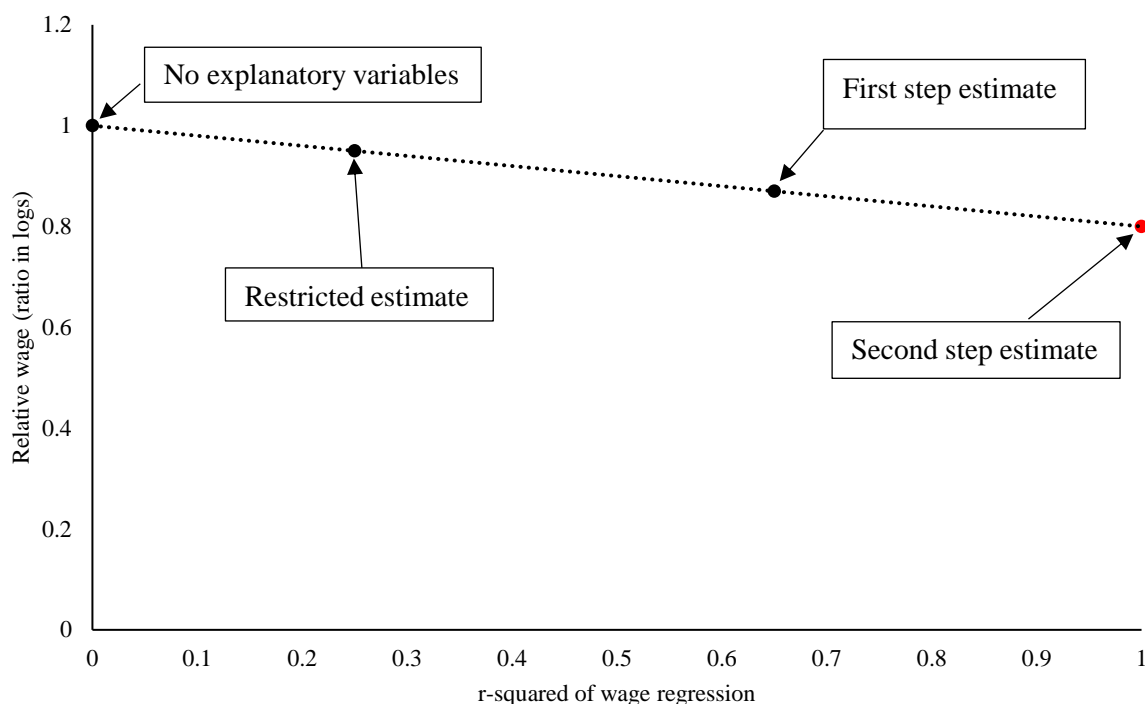
<sup>24</sup> This is admittedly a crude approximation, and one that is hard to quantify. The accuracy of the approximation can vary, and indeed varies, across country as well as which variables are excluded from the nested model.

<sup>25</sup> The approach also relies on the similarity of the data generating process for wages of employees and notional wages of the self-employed, but this will be true of any approach that uses the former variable as the latter is unobservable.

<sup>26</sup> As opposed to the approach of just not adopting any correction mechanism.

<sup>27</sup> The correction mechanism would greatly benefit from further research. These efforts could be focused on two paths. The first would be finding a more general solution that is more robust to changes in the specification, this analysis should be carried out in a simulation framework. The second would be exploiting the information that can be retrieved from self-employment income. Even if overall income only partially reflects labour income, and is subject to measurement challenges, the variable could provide a useful proxy as a second best pseudo out-of-sample exercise.

Figure 1. Illustration of the log measure of the relative wage for own-account workers



Note: The no explanatory variables scenario would be equivalent to simply assuming the same average wage for employees and own-account workers (thus the ratio of the logarithms of the averages is 1 and the r-squared is 0). The restricted estimate,  $(\delta_{OAW}^R)$ , corresponds to a modified first step estimate with some of the explanatory variables discarded. The first step estimate,  $(\delta_{OAW})$ , is produced using the full set of explanatory variables. Finally, the second step estimate,  $(\theta_{OAW})$ , is produced by applying the correction formula to the first step estimate. The values are calibrated to guarantee that the restricted estimate falls in the line, joining the no explanatory variable case and the first estimate for illustration purposes.

With the described methodology, the relative wage of each group of the self-employed can be computed directly from the estimates in levels. These relative estimates are labelled  $\gamma_{OAW}$ ,  $\gamma_{CFW}$ ,  $\gamma_{ERS}$  for own-account workers, contributing family workers and employers. It must be highlighted that a manual selection step, identifying and discarding breaks and outliers, is also carried out to produce consistent and comparable estimates. Table 4 in the appendix details the surveys used. Similarly, for a few surveys, extremely high salaries have been detected, probably a consequence of coding errors during data collection. In these cases, rather than remove the whole survey, the outliers are winsorized by taking the value of the 99.5 percentile. In a few other cases, the outliers are coded to indicate missing values, thus those observations are removed from the sample.<sup>28</sup>

#### 4. FROM THE UNADJUSTED LABOUR INCOME SHARE TO THE ADJUSTED SHARE AND THE GROSS CAPITAL SHARE

The measure of the adjusted labour income share proposed is based on two main principles. First, any adjustment to account for the self-employed has to take into account the heterogeneity within this group. In order to achieve this, detailed data on status in employment, which subdivides self-employment into

<sup>28</sup> The former approach is used for Ghana 2013, Gambia 2012, Namibia 2012-2016 and Zambia 2017. The latter approach affects the PNAD survey of Brazil, Côte d'Ivoire 2012, 2013, Cameroon 2014, and Cape Verde 2015, and Niger 2011.

three different groups: own-account workers, contributing family workers, and employers is used. Second, the relative wage of each group of the self-employed (necessary to adjust the compensation of employees) must be informed by empirical evidence. To do this the relative wages imputed are based on the microdata analysis outlined above.

The proposed adjusted labour income share is:

$$ILO LIS = \frac{Compensation\ of\ Employees}{GDP} \cdot \frac{\% Employees + \gamma_{OAW} \cdot \% OAW + \gamma_{CFW} \cdot \% CFW + \gamma_{ERS} \cdot \% ERS}{\% Employees} \quad (13)$$

The correction is clearly related to the variant based on self-employment data (G3 variants) described previously, but it enriches the information set by using more detailed status in employment data as well as estimated – as opposed to assumed – relative wages for the self-employed. Therefore, the correction represents a departure from the rule of thumb approach commonly used, and instead is based on empirical analysis of relevant data. The data of status in employment, whether a worker is an employee, an employer, an own-account worker or a contributing family worker, is obtained from ILO (2018a).<sup>29</sup> The data on compensation of employees and GDP is derived from the international repository made available online by the United Nations Statistics Division.<sup>30</sup> Finally, notice that the gross capital share can be computed as one minus the adjusted labour income share.

## 5. FROM THE LABOUR CAPITAL SPLIT TO THE COMPLETE LABOUR INCOME DISTRIBUTION

The data used to compute the labour income share presents an interest in itself. In particular, for every worker in each microdataset their actual labour income is available, or an imputed one for the self-employed and employees with missing data. This distribution is used to compute the labour income distribution, the distribution of the income that does not accrue to capital or depreciation. To the best of the author's knowledge, this is the first attempt to estimate an international panel dataset of the labour income distribution. Nonetheless, this effort is related with the empirical income distribution literature, in particular with the poverty estimation and distributional national accounts fields. The production of the distribution of labour income relates to two topics commonly discussed in the literature that need to be addressed. The first is the convenience of pinning distributional data to a national accounts magnitude such as GDP. The second is the suitability of using survey data to compute inequality metrics. In this section it is argued that focusing on labour income, and using as a source wage data, has a substantial impact vis-à-vis these two issues.

The contentious topic of matching distributional data to national accounts data, or whether to directly consider the survey mean as the valid average income, has its greatest spotlight in the poverty literature. Both approaches have been used extensively. Deininger and Squire (1996) pioneered in the combination of within country survey distributions with GDP to produce international distribution estimates. Sala-i-Martin (2006) also follows this approach. On the other hand, studies like Milanovic (2002, 2005, 2015) and Chen and Ravallion (2010) scale their distributional estimates on the basis of the same survey mean. Deaton (2005) discusses at length the implications of each methodology. The divergence of

<sup>29</sup> Importantly the source includes both observed and modelled data. The metadata of the results of the present study does clearly identify whether the input data for status in employment is modelled or observed.

<sup>30</sup> In a few cases there are breaks in the provided series due to reported methodological changes or to unspecified causes. In case of a reported methodological change the series is backtracked using the most recent methodology as the baseline level (using the average ratio of the two series during the time period of overlap as the correction factor). When a clear break or an outlier is detected in the series with no further information, the observation in question is discarded.



methodologies has not occurred without a lively debate. Pinkovskiy and Sala-i-Martin (2016) use night-time lighting data to assess the relative error of survey means vs GDP, and find strong evidence of much smaller errors on the latter. In contrast, Anand and Segal (2008) argue forcefully against pinning survey distributions to national accounts magnitudes. The case made in the latter study is that using GDP to anchor survey distributions is misleading, given that the magnitude includes depreciation, retained earnings, or government revenue that is not redistributed. Even if a sub-item of GDP is used in the context of consumption survey data, such as households and non-profit consumption, conceptual problems and inconsistencies remain, as discussed by Deaton (2005).

The current analysis focuses only on labour income, and exclusively uses wage data (more precisely labour-related earnings of the employees). Does this alter the perspective about the convenience of pinning distributional survey data to national accounts? It does, and in an important way. Whereas there is no consensus in the income distribution literature on the issue, when focusing on labour income, the use of the national accounts item “compensation of employees” (after having it properly corrected for self-employment)<sup>31</sup> as an anchor for labour income distributions should not be contentious. This circumstance should not be taken as an argument for or against the use of national accounts to anchor total income distributions. The following arguments to use national accounts data as an anchor only concern labour income. To put it simply, the debate that has been ongoing for decades with respect to overall income is outside the scope of the present study.

There is a clear consensus in the empirical literature about using GDP and compensation of employees as a basis to compute the share of labour income. All of the studies – to the best of the author’s knowledge – that aim to estimate the labour income share use national accounts data. This is due to the match between national accounts and the concept to be estimated. For instance “compensation of employees” is an uncontroversial measure of the labour income of employees.<sup>32</sup> There is a similar consensus that the denominator of the labour income share should be some aggregate measure of the overall economy (which can be GDP, GDP minus depreciation or other related measures). At the same time, the exclusive use of household survey data is not sufficient to split income into labour and capital, since the non-household sector can have income, and the national accounts framework provides tools to address under-coverage of income. The distribution of the labour income share amongst workers has not been generally studied, so it is not possible to claim a similar consensus. Nonetheless, many of the same arguments used in the case of the labour income share remain identical for labour income distribution. Compensation of employees does adequately capture the aggregate labour income of employees, therefore using it as the anchoring mean of the distribution of employee income can be easily defended. Extending this to the total labour income of the workforce, via the self-employment adjustment, does not alter this circumstance.

The second topic of the empirical literature on income distribution raised above is the convenience of using enterprise survey data or administrative data (including tax data) over household surveys. Piketty, Saez and Zucman (2018) argue that tax data present an advantage over household survey data to compute the income distribution mainly due to under coverage of high incomes in surveys. The recommendation of adjusting survey data with administrative sources, or to use administrative sources directly, is forcefully made in several studies of the World Inequality Lab initiative: Blanchet, Flores and Morgan (2018), Alvaredo, Chancel, Piketty, Saez, Zucman (2017), Alvaredo, Atkinson, Chancel, Piketty, Saez and Zucman (2016).

In this study, household survey data are not adjusted with administrative or establishment survey data. The first reason is overwhelmingly practical: the data are not readily available for a substantial number of countries. Furthermore, metadata availability is as binding as data restrictions. To make a proper comparison between two data sources, a necessary step before adjustment, very precise metadata are

---

<sup>31</sup> Or alternatively, using it uncorrected to produce only the distribution of labour income amongst employees.

<sup>32</sup> Albeit certain discrepancies might arise due to the study of tax incidence.

required to ensure that the same coverage, omissions, and concepts are used across sources. Thus, the study of systematic differences and suitable adjustments based on additional sources is left for future research.<sup>33</sup> The second reason is that focusing on labour income represents an important change in the circumstances. Thus, it is worth considering if the adjustment is required for wage data, as it is the case for overall income. Furthermore, the bias outlined in the above studies, a clear underestimation of the income of high earners, must also be re-examined in the case of wage data.

The complete answer to both questions can only be given when administrative data and/or establishment survey data and metadata are systematically obtained, homogenized, and cross-checked against labour survey data. Unfortunately, at this time data restrictions do not allow for such an exercise. There are, however, several reasons that indicate that the adjustment of household survey data is not a requirement to study labour income. This should not be taken as a claim that either administrative data or establishment surveys do not capture labour income well. Both types of data can be considered as an excellent source for wage data. Tax evasion is certainly a concern for both types of sources, nonetheless household survey data have often been shown to be affected by it as well, Cabral and Norman (2018). The issue is simply if labour survey data are accurate enough to measure the wage distribution. The answer to this question, based on existing evidence from literature as well as additional sources considered in the present study, seems to be yes. How can wage data from household surveys be more accurate than overall income from the same survey? It must be highlighted that top incomes are often found to have large shares of capital income, more so than the average. For instance, Piketty, Saez and Zucman (2018) estimate that the top ten per cent earns almost half of its income through capital, whereas for the bottom 90 per cent the share is below 20 per cent. Therefore it is possible that the underestimation of top incomes from household surveys, which seems to be an important cause for biased distributions, is for the most part a consequence of underestimating capital income, instead of labour income. Encouragingly, this seems to be the case.

Evidence is presented in favour of the hypothesis that using household survey data to estimate labour income is substantially less problematic than overall income, and that reasonable estimates of its distribution can be obtained. The first piece of evidence analysed is a country case study of Belgium in the year 2014. For that year both the household survey, European Union Statistics on Income and Living Conditions (EUSILC), as well as an establishment survey, the Structure of Earnings Survey (SES) are made accessible by Eurostat. The Belgium SES is an excellent benchmark to test the validity of labour survey data. First, the SES gross annual earnings in the reference year are derived from the National Office for Social Security, and hence they do not only derive from an enterprise survey, they reflect administrative tax data. Second, the SES is based on a stratified random sample of local units of companies. In a second stage, a random sub-sample of the employees of each selected local unit is drawn. Third, the sample size is substantially bigger than the EUSILC one. These three elements rule out typical household survey problems with respect to income estimation, such as top-coding errors, small samples, or selection bias in the sample. Unsurprisingly, the Belgium SES 2014 data are found to be broadly consistent with national accounts data. Therefore, this source can be considered an excellent benchmark for a case study on the validity of labour income derived from labour surveys. The comparison between data from different sources has to be done with extreme care, particularly when comparing income distribution data at the percentile level. Differences in coverage and methodology can result in drastic differences in labour income distribution that are only a statistical artefact, and not a quality evaluation of labour surveys. The particular coverage of economic sectors and activities, and the match between the reference period of the employment definition, and the reference period of income measurement are crucial to homogenize in order to compare the two sources. The Belgium 2014 SES excludes companies with less than ten employees, as well as employees in the agriculture, public administration, and extraterritorial organisations activities. The data of the EUSILC is adjusted to

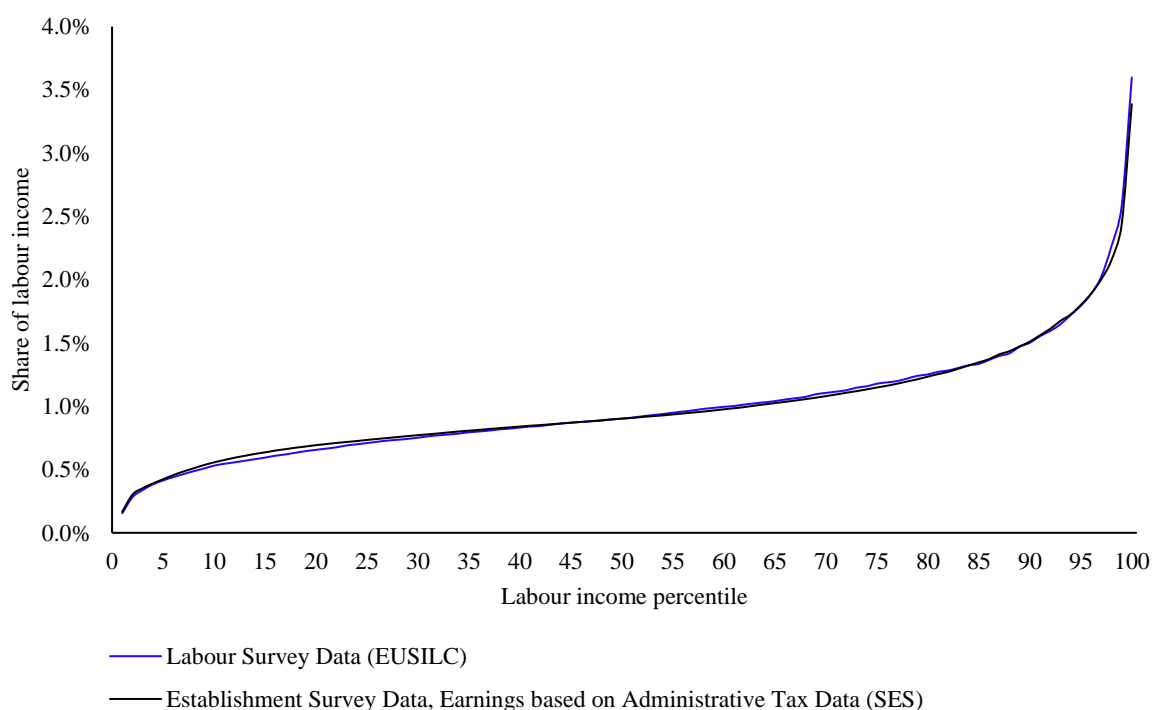
---

<sup>33</sup> At the time of writing the process of obtaining a substantial number of establishment survey data is underway. This will allow to check the robustness of the current methodology.

exclude these categories.<sup>34</sup> Additionally, given the different recording practices of SES and EUSILC, the comparison is restricted to employees in both sources that worked during the whole reference year. After taking these steps to ensure broad consistency between the sources, the distribution by percentiles of gross annual earnings can be produced for each data source. The underlying SES income concept corresponds to item D11 of the European System of National Accounts, and hence excludes the employers' contributions on behalf of the employee, accordingly the same definition is used for the EUSILC.

The results of the exercise can be found in Figure 2, and they are very encouraging as the estimated distributions are quite close. The result is remarkable given that the homogenisation of the two data sources has not been exhaustive<sup>35</sup> and that the homogenisation steps already taken cause shifts on the distribution much larger than the existing SES-EUSILC discrepancies. Furthermore, the EUSILC distribution is produced from 2'944 observations, whereas the SES one uses 105'789. This exercise, even if it based on a single country and year, highlights that household surveys can capture the labour income distribution with a reasonable degree of error. The exercise also makes clear that in order for the cross check of household surveys against other labour income sources to be informative, it must be carried out at the distribution level (not with simple summary statistics such as the coverage rate) and with a data and metadata intensive procedure of homogenisation.

Figure 2. Labour income distribution, comparison between household survey and establishment survey (based on administrative tax data), Belgium 2014



Note: Comparison of annual gross earnings distribution, excluding employers' social contributions. The two data sources have been partly homogenized in terms of economic activity, firm size, and work during the reference period. The results illustrate the comparability of labour income from household surveys with other sources

<sup>34</sup> Given the economic activity classification available all the categories corresponding to NACE v2 A, R, S, T, U in EUSILC and R, S in SES are excluded from the exercise.

<sup>35</sup> Even with all these homogenisation procedures, the resulting weighted sample presents substantial differences in the total number of employees covered. Further research is needed to pin down exactly the factors. A potential source for the discrepancy would be employees that have changed of work (and would be included in the EUSILC but not the SES). This discrepancy implies that the coverage rate cannot be interpreted straightforwardly.

typically considered of higher quality. The comparison is made between the household survey EUSILC and the enterprise survey SES, which bases earnings data on a subsample of the social security administrative register.

Additional evidence in favour of the hypothesis that labour surveys provide reasonable estimates of labour income is provided by the European Union statistical agency. Eurostat publishes experimental results concerning EUSILC microdata coverage rates of national accounts data. For labour income, the benchmarking is done between the microdata gross wages and the national accounts wages and salaries. When comparing microdata sources against national accounts it is crucial to use the appropriate benchmark. In EUSILC data the employers contributions are only experimentally collected (and are not present before 2007), therefore it is important to subtract those from the benchmark. Mismatch between the precise concepts that are being compared can greatly overstate the actual discrepancy. Therefore, given the large heterogeneity in labour income household survey data, benchmarking it against national accounts has to be done on a case by case basis. When comparing the two data sources, between 2008 and 2015 the median coverage rate for the 28 European Union members of the EUSILC data was 94.2 per cent (and an interquartile range of 88.1 per cent to 99.2 per cent), suggesting a reasonable coverage and hence a low potential for bias in the microdata.

A study of 19 OECD countries, Fesseau, Wolff and Mattonetti (2013), points to a similar conclusion. Comparing household surveys against national accounts aggregates, the results show that on average the micro source accounts for 93 per cent of the macro wage income (with a range of 71-107 per cent). Rothbaum (2014) finds that in the US the Current Population Survey matches between 96 and 99 per cent of national accounts wage income. Does this finding contradict the notion of under-coverage of high earners from Piketty, Saez and Zucman (2018)? Not at all, Fesseau, Wolff and Mattonetti (2013) show that capital income, such as interest and distributed income by corporations from micro sources only covers on average 53 per cent of the national accounts concept. Therefore, the underestimation of income for high earners seems to be an overwhelmingly capital income-related phenomenon.

Admittedly, the average coverage rate of 93 per cent shows that there are certain discrepancies. Furthermore, the coverage in one of the countries is as low as 70 per cent. The mismatch nonetheless does not necessarily indicate that there will be an under-estimation of labour income at the top. Whereas the average 93 per cent could suggest that the missing seven per cent should be assigned to top earners, Fesseau, Wolff and Mattonetti (2013) report that the employers' social contributions from micro sources account for 83 per cent of the national accounts concept. Given that social contributions usually have a cap, this could lead to an overestimation of labour income accruing to top shares. Therefore a systematic cross-check of labour survey data with establishment surveys or administrative registers is indeed a highly desirable practice. The lack of the necessary data is the only reason to not do so. Nonetheless, the available empirical evidence suggests that the adjustment of household survey data to analyse labour income is much less relevant than for overall income and that the direction of the bias is unclear. Thus, it seems that this issue should be addressed in future research efforts.

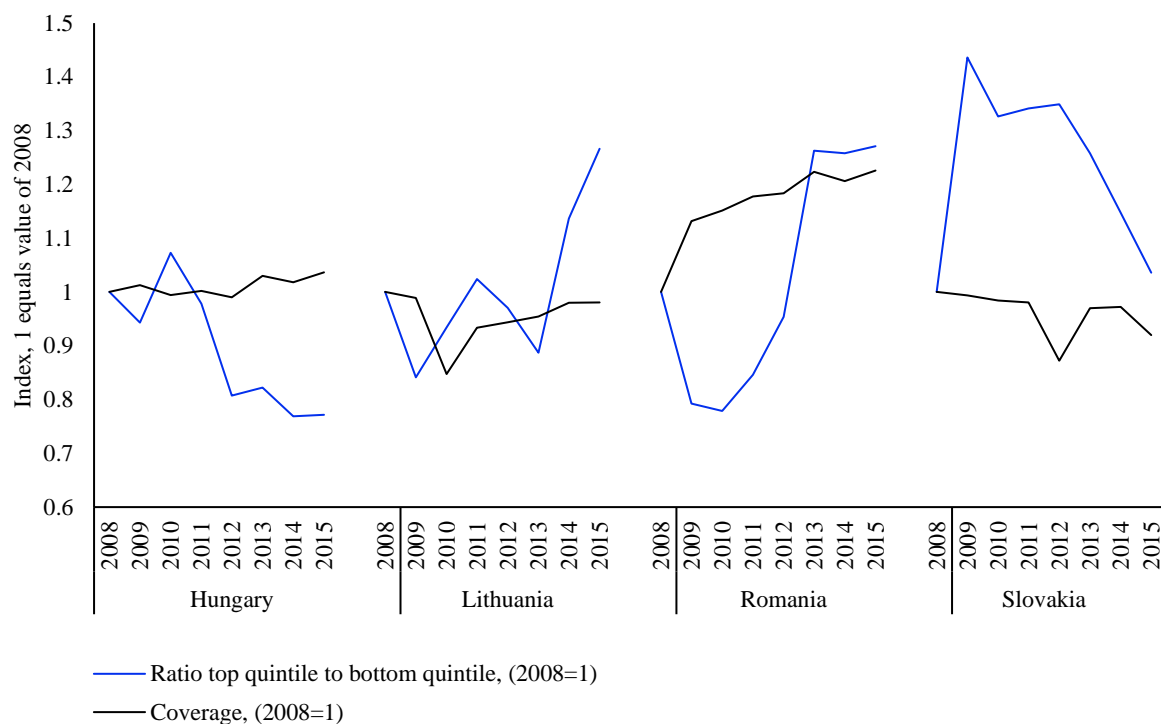
Furthermore, as discussed above in the Belgium 2014 comparison between labour surveys and establishment surveys, mismatches in the coverage rates can easily derive from different samples of the population and methodologies, without substantially affecting the estimated labour income distribution. The microdata coverage rates thus are a useful tool to analyse the potential for bias in the data, nonetheless the bias itself should be analysed. Concerning labour income there does not seem to be a clear relationship between coverage rates and inequality over- or underestimation. For instance, using a fixed effect regression of the EUSILC data coverage rate against the ratio between the top and bottom quintile, no significant association is found.<sup>36</sup> To illustrate this point, the evolution of the coverage rate and the ratio between top and bottom quintiles is plotted in Figure 3, for EUSILC countries with coverage gaps below 80 per cent (for at least one year). As can be observed, no systematic relationship

---

<sup>36</sup> The corresponding t statistic is 0.23, suggesting a very wide confidence interval practically centred at 0.

appears to exist between the evolution of coverage and inequality. This suggests that the under coverage of labour income from microdata when compared to national accounts does not derive from a systematic error related to inequality.

Figure 3. Evolution of inequality (measured as the top to bottom quintile ratio) and coverage rate, selected EUSILC countries

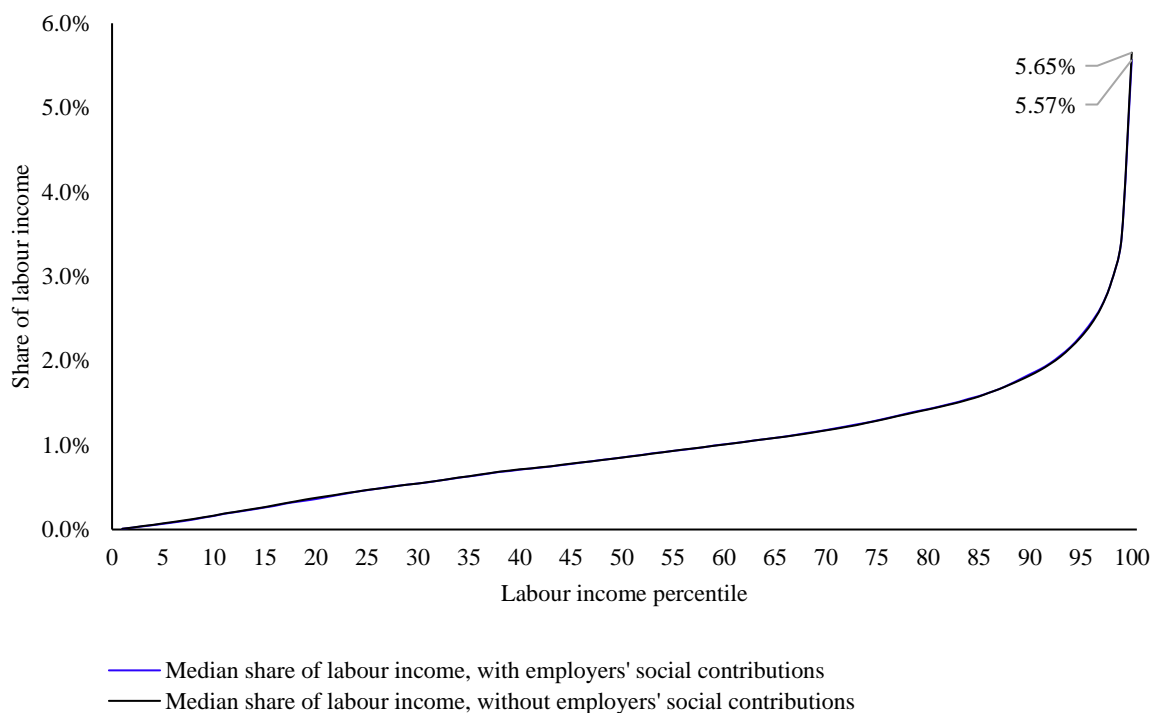


Note: Both the coverage rate and the inequality measure are set to one for the year 2008. The selection of countries is done on a basis of having the coverage rate below 80 per cent for at least one year. The figures for four countries highlight that changes in coverage rates are not associated with changes in the inequality measure. This is consistent with the regression results carried out for the whole EUSILC sample.

These encouraging results highlight the importance of focusing not on coverage rates but on the effects that these have on the income distribution. Depending on the source of the under coverage, the impact on the distribution can be quite low even for substantial coverage gaps. For instance, one important source of under coverage of labour income from household survey data has to do with taxes and employers' social contributions. Indeed, even if the ICLS standards require that labour income in household surveys includes gross wages and employers' social contributions, this is not always the case. Given the magnitude of these items in the overall compensation, their omission can cause large drops in coverage rate, of the order of 30 per cent or more. Nonetheless, this under coverage is not necessarily related to a markedly biased estimate of the labour income distribution. To show this, the employee labour income distribution is estimated with and without employers' social contributions, based on EUSILC microdata – since the data source allows to separate the item. Excluding the employers' contributions has a large effect on the coverage rate, the median employee income is 16.8 per cent lower when excluding the item. This implies that even if we started with a median coverage rate of 100 per cent, excluding the employers' contributions would bring down this coverage to 83.2 per cent. Nonetheless, the effect on the median income distribution is quite small, as can be seen in Figure 4. In fact the largest effect (which can be observed for the top labour income percentile) on the estimated median distribution is below 0.1 percentage points. The disparity between the change in median

coverage rates and the stability of the median distribution highlights the importance of benchmarking the whole labour income distribution, and not placing an excessive focus on the total sum.

Figure 4. The effect of omitting employers' contributions on the labour income distribution



Note: Not including employers' contributions causes a median decrease, across all EUSILC countries and years, of 16.8 per cent in total labour income. However, when analysing the median share of labour income by percentile (across all EUSILC countries and years) the effect is quite small. The figure illustrates this by showing the median share of each labour income percentile with and without employers' social contributions.

The data used present several limitations beyond the type of data source. First, the surveys do not always follow ICLS standards, so in some cases the labour income variable will present methodological deviations. Second, only data on the main job<sup>37</sup> is used to compute labour income. The decision to exclude additional jobs is due to a homogenisation purpose, as many surveys do not have labour income for additional jobs, as well as a simplification of the imputation procedure of labour income for the self-employed. Additionally, labour income from additional jobs as an employee tends to be very low in comparison to the income derived from the main job (the median is 1.7 per cent). Furthermore, when included it produces a very small effect on the estimated labour income distribution of employees. Third, the labour income recorded is nominal, not real. Thus to the extent that price differences are positively correlated with wage differences (which is certainly likely), nominal inequality will overestimate real inequality within a country. These limitations notwithstanding, the estimates of labour income distribution and inequality represent a novel data collection that has interest by itself, beyond the scope of the labour income share.

Taking as given the labour income observations from the modelled microdatasets, the production of distributional measures is straightforward. Importantly, given that the benchmark of employment follows the ICLS standards, the labour income is estimated on a per worker basis, not on a full time equivalent basis. Afterwards, the labour income data at the micro level and the percentiles are

<sup>37</sup> EUSILC data, used in many European countries, as well as data from the Luxembourg Income Study, only allow to identify gross labour related income for all jobs (as an employee). Consequently, this magnitude is the one used.

computed.<sup>38</sup> Thus, one can know the *n* per cent of income that goes to the bottom one per cent, the second percentile, and so on. The level of labour income of each percentile can be easily obtained by multiplying the share by the overall labour income (which can be computed from GDP in PPP\$). There are only two methodologically noteworthy issues.

Computing the percentiles requires having the observations grouped into quantiles representing an identical mass. Given that the microdata is composed of discrete individuals, it is simply not possible to have exactly the same mass in each. In this case the marginal observation that is “between” two percentiles of labour income has to be divided between the two. For instance, let’s imagine that a worker that earns US\$1000 is located, in the cumulative distribution of labour income (of workers ranked by labour income), between 0.9 per cent and 1.1 per cent. In this case the split is done proportionally adding US\$500 to the first percentile and US\$500 to the second percentile. Given that the sample size of the labour survey data tend to be large,<sup>39</sup> this approximation works very well in practice.<sup>40</sup>

The other methodological issue concerns imputed labour income. Recall that for employees with no labour compensation data and for self-employed workers a labour income has been estimated via a regression model. If unadjusted, the labour income distribution would greatly underestimate the true variance. The reason is that the fitted values of a regression do not include stochastic variation. In fact, the residual of the regression – which is the estimated noise – is set to 0 when predicting values. To avoid this shrinkage of the imputed distribution of labour income, the residual is added back to the estimates. The estimated residuals of the workers with earnings data are randomly assigned to the imputed observations.<sup>41</sup> In this manner, the average imputed labour income is preserved, and the variance of the original wage data is recovered.

## **6. FROM THE LABOUR INCOME SHARE AND DISTRIBUTION AT THE COUNTRY LEVEL TO THE GLOBAL AND REGIONAL LEVEL**

The methodology described until now has been focused on how to produce estimates of the labour income share and distribution for countries with labour survey data available, and in particular when access to the full microdataset is possible. Thanks to the use of international repositories, mainly from ILO and the United Nations Statistics Division (UNSD), it is possible to carry out this exercise for a large number of observations. In Table 1 a summary of the availability – after data cleaning – can be found. The coverage of individual variables is relatively rich. The notional maximum sample is roughly 2’650 observations, 14 years x 189 countries. Given the inclusion of data scarce regions, such as Africa and Asia, a full panel data set is an ambitious goal. The most common data source is national accounts data on compensation of employees and GDP (the components for the unadjusted LIS), covering 46 per cent of the panel. Status in employment data, from ILO (2018a), also presents coverage of 46 per cent. The availability of labour survey data is roughly half, depending on the variable of interest, ranging from 20 to 22 per cent. Taking into account the greater level of detail of microdata, compared to aggregate data, this represents a comparatively high coverage rate.<sup>42</sup> Nonetheless, both the coverage rate of aggregate data as well as microdata are quite low, and point to data scarcity challenges in the

---

<sup>38</sup> The estimates are disseminated at the percentile level. Nonetheless, it must be highlighted that often it is better to work at a higher level of aggregation, since the estimates are noisy.

<sup>39</sup> The smallest sample of workers consists of several thousand and many countries survey hundreds of thousands.

<sup>40</sup> In many cases the salary of the marginal worker is exactly the same than the one of the worker below them in terms of labour income, as well as the one above, so the approximation is exact.

<sup>41</sup> If there are more workers that require imputation than there are observations of earnings, the residual is reused as many times as needed.

<sup>42</sup> This relatively high coverage is a direct consequence of the efforts to provide open data, or at least restricted-use accessible data, by the National Statistical Offices as well as the harmonizing initiatives like the ILO Harmonized Microdata collection.

global arena. This scarcity can be easily grasped when considering only the results that have actual observations on all the magnitudes of interest, in this case the global coverage drops to 12 per cent.

Table 1. Data availability by type of variable

Variable estimated/used	Observations	Percentage
Relative wage: employers	567	21%
Relative wage: own-account workers	583	22%
Relative wage: contributing family workers	534	20%
Labour income distribution	548	21%
Compensation of employees & GDP	1 224	46%
Status in employment	1 207	46%
<b>All of the above combined</b>	<b>329</b>	<b>12%</b>

Note: Coverage of variables of interest: total number and percentage of notional maximum, 2004-2017.

This magnitude makes clear that there are substantial gains in imputing missing data. In the first place, some of the missing variables have very little impact on the final result.<sup>43</sup> Similarly, both across the time series and across the variables, there are important correlations that can be exploited to impute a reasonable set of values for the missing observations. Finally, one of the objectives of the present study is to analyse global and regional trends related to labour income and its distribution. As described in Crespi (2004), when analysing datasets with missing observations, non-response bias has to be considered. When working at the country-year level, non-response can hardly be considered as randomly assigned. A clear relationship exists between economic development and data availability. Biases across time are also present, with both the most distant and recent years being under-represented. Therefore, the objective of analysing trends at both the global and regional level makes it imperative to account for the non-response bias. One practical approach to deal with non-response bias, when producing global and regional aggregates, is first to impute the observations that are missing from the sample and afterwards compute the aggregates combining the actual and imputed observations.<sup>44</sup>

The methodological approach to imputation draws heavily from the approaches used in Crespi (2004), ILO (2017) and ILO (2018a). There are three main types of variables that are imputed for this study: the relative wages of the self-employed, the distribution of labour income, and the adjusted labour income share.<sup>45</sup> The relative wages of the self-employed and the distribution of labour income are estimated in the following ways depending on the data availability. In the first place, linear interpolation is used to impute interior points of a time series. This can be carried out in countries where several data points are present, but not contiguous. The performance of this simple method is satisfactory, mainly due to the persistence and trending behaviour of the underlying variables. The second method is a country fixed effects imputation based on a least absolute deviations (LAD) regression,<sup>46</sup> including two explanatory variables: the (log) share of employees and (trend) GDP per capita. This method can be

<sup>43</sup> A straightforward example is data from European countries. In many cases, the contributing family workers' relative wage could not be imputed given the very low rates of this employment category. It is clearly not an acceptable outcome to lose observations due to not having a magnitude that was not imputed precisely because the incidence was too low to compute – hence, it would present a very limited impact on the actual data.

<sup>44</sup> The imputation of missing observations has an advantage over modelling the aggregates directly (weighted by probability of non-response): the estimates are consistent across different levels of aggregation.

<sup>45</sup> The data on status in employment is already imputed when missing, ILO (2018a).

<sup>46</sup> In this study, all imputation exercises use least absolute deviations instead of the more common ordinary least squares. LAD regression is chosen for its lower variation of estimates in response to repeated sampling tests. Each regression is weighted to reflect the differential response rates across income, region and year. Furthermore, the regression results are adjusted to account for the difference between the predicted and observed values. Thus, a correction ratio is computed, by dividing the observed value by the fitted value, and then adjusting the fitted value accordingly. This is done to avoid any discrepancy between the fitted value and the observed value, when it is possible to establish this comparison. For consistency, the predicted values without an observable counterpart are adjusted by the interpolation or extrapolation of the correction ratio.



used in countries with at least one observation of the variables of interest. Finally, for countries without any observation, a LAD regression is carried out with region fixed effects and the same variables as in the previous case.

The adjusted labour income share is only imputed in a subsidiary basis. So, if data on unadjusted labour income share is present, it is combined with the imputed or observed relative wage following equation (10).<sup>47</sup> When no such data is available, a similar strategy to the previously outlined is followed.<sup>48</sup> If at least one observation at the country level is present, a country fixed effects regression is run, otherwise regional and income group fixed effects are used. In this case the variables included are the (log) share of employees, standardised GDP growth, and the relative wages of the self-employed. The relative wages are included for two reasons. First, there are actual<sup>49</sup> observations of relative wages that do not overlap with unadjusted LIS and hence they add new information. The second reason is that, even if they are imputed, their inclusion helps decouple the dual effect of share of employees. Low shares of employees tend to mean higher adjustment factors but lower unadjusted LIS. Finally, the inclusion of a cyclical indicator of GDP, instead of a trend measure, reflects the fact that adjusted LIS is not strongly related to level of GDP per capita, but it is quite (counter) cyclical.

Once the imputation of missing countries is completed, computing global and regional aggregates is just a matter of adding the appropriate magnitudes across countries. Nonetheless, for distributional data a final difficulty must be overcome. Population, and hence employment, across countries is highly heterogeneous. Thus, for a given quantile at the aggregate level a single country quantile can account for a very large share of its mass. For instance, say that we want to split the global workers in deciles using country level deciles. Given that a decile from India or China can account for 15 per cent of a global decile, the direct production of the global deciles is not possible.<sup>50</sup> To deal with this issue two complementary strategies are adopted. First, the quantile adjustment procedure that was used in microdata (and outlined in the previous section) is used to split the mass between quantiles with uneven mass. Second, global and regional aggregates are computed only at the decile level (and not at the percentile one). The combination of the two strategies is necessary, given that heterogeneity in level of employment is very large even at the global level, but is exacerbated further in several regional groupings. Nonetheless, a reasonable degree of aggregation error can be obtained using deciles at the aggregate level (produced by percentiles at the country level) and applying the correction method.

---

<sup>47</sup> The use of imputed coefficients combined with actual unadjusted labour income share data can be interpreted as a restricted model. In particular, it restricts how the relative wages affect the adjusted labour income share. In unreported robustness exercises, this approach has been shown to present both lower volatility and higher accuracy than the direct imputation of the adjusted labour income share.

<sup>48</sup> The linear interpolation is not carried out, as it is only applicable to a very small number of observations.

<sup>49</sup> In the sense of microdata estimated relative wages.

<sup>50</sup> Certain deciles would have substantially more than ten per cent of the mass, whereas others would have less.

### 3. RESULTS

---

This section presents and discusses the results. The modelled estimates include the adjusted labour income share, detailed distributional data of labour income (at the percentile level) and imputed relative average wages by detailed self-employment group for 189 countries. Additionally, the adjusted labour income share and distribution of labour income at the decile level are available for global and regional aggregates. All these data, together with the corresponding metadata that clearly identify how each observation is produced, are available on ILOSTAT ([ilostat.ilo.org](http://ilostat.ilo.org)). Given the magnitude of the produced dataset, the present discussion does not aim to be exhaustive, rather it aims to highlight key results and relate them to methodological practices.

#### 1. THE (ADJUSTED) LABOUR INCOME SHARE

The global adjusted labour income share presents a substantial downward trend in the period from 2004 to 2017 (Figure 5). The decrease was temporarily reversed during the financial crisis in 2008-2009, due to a clear countercyclical behaviour. Both Europe and the Americas are key drivers of the global decline in the labour share. Since 2004, the share in the Americas declined by 1.6 percentage points, and in Europe by more than 2 percentage points. In the Americas, the United States, Brazil, Mexico and Canada are by far the largest economies. The labour share in the United States lost almost three percentage points between 2004 and 2016.<sup>51</sup> In contrast, Brazil has presented an upward trend of a greater magnitude. Mexico presents a decline larger than the one in the United States, whereas Canada's labour income share has remained stable. In Europe, despite the counter-cyclical increases for the 2008-2011 period, Germany, the United Kingdom, Italy and Spain present sizeable net declines between 2004 and 2016.

The pattern of long-run decreases, with countercyclical behaviour, does not hold in general outside of developed countries. Asia does present an important decline, with a countercyclical blip in 2008-2009, mainly due to the behaviour of the labour share in India. Africa, on the other hand seems to have been decoupled from the global declines, and since 2010 its labour income share is steadily raising (albeit from the lowest level of the considered regions). It is important to highlight that in the last two regions the data availability is scarce, and hence the regional figures present a large degree of uncertainty.<sup>52</sup>

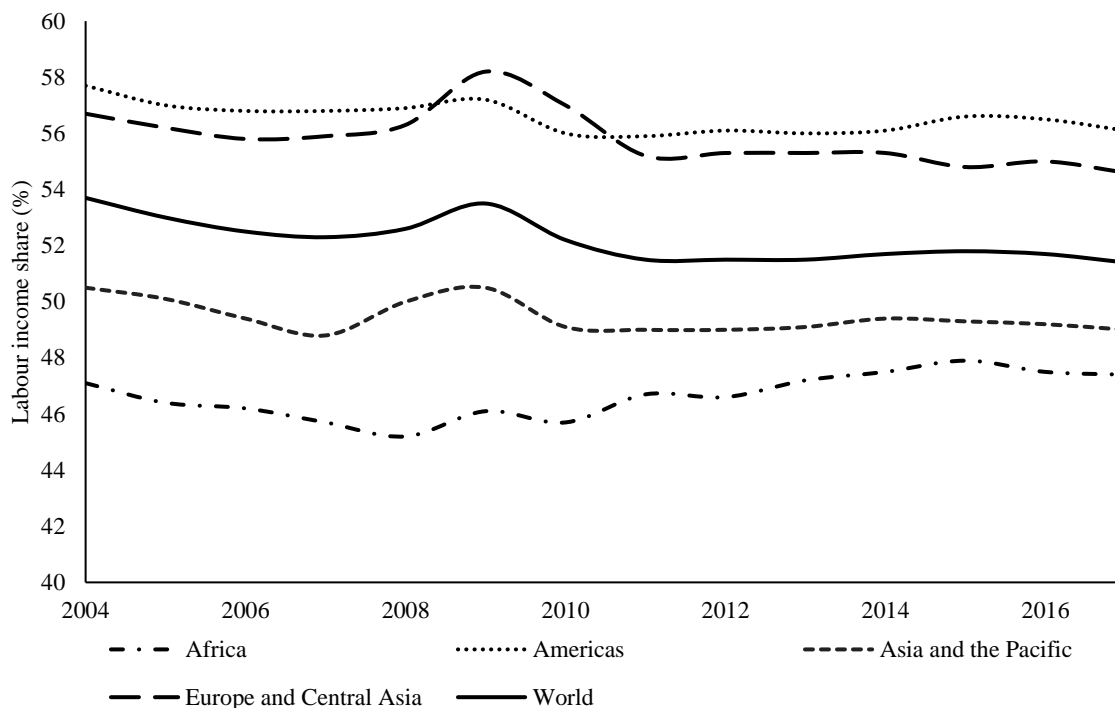
Aside from the underlying trends, Figure 5 shows that the regional labour income shares lay within a relatively narrow range. This low dispersion is a consequence of the adjustment. The unadjusted, or "raw" labour income share presents a strong correlation with income. The adjustment lessens this degree of correlation because economic development is strongly associated with lower levels of self-employment. Nevertheless, a regional pattern related to income can still be observed. The higher-income regions present adjusted labour income shares above the global level, and vice versa.

---

<sup>51</sup> The national accounts data from the UNSD repository ends in 2016 for many countries. Therefore, often 2016 is referred to as the end of the sample for country level data. The model does indeed impute the missing 2017 observations, and at the aggregate level they are shown. These estimates however have to be met with care since they are subject to greater uncertainty.

<sup>52</sup> In India, the last microdata set used corresponds to 2010 and since then the relative wages must be imputed. Furthermore, China presents very particular data characteristics, and as explained in the appendix and ad-hoc procedure is used to take them into account. Africa, presents very sparse sampling, resulting in uneven coverage across years. Finally, the Arab States regional results are not presented, given the degree of data limitations. Two elements are of particular concern: the almost absolute lack of microdata from the region and the statistical measurement of foreign workers.

Figure 5. Global and regional adjusted labour income shares, 2004-2017

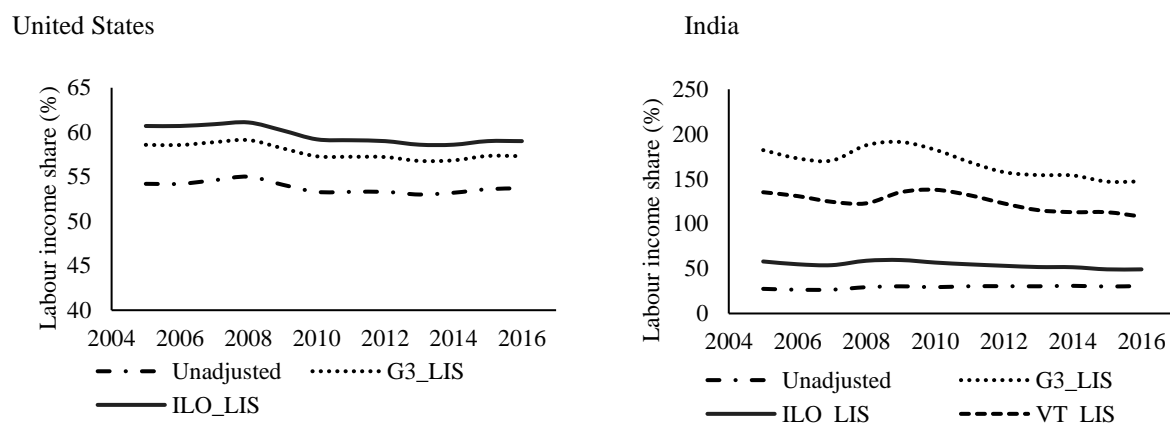


In order to analyse the effect of the microdata-based adjustment, as opposed to commonly used rules of thumb, it is illustrative to consider two very different countries: India and the United States. Figure 6 presents the unadjusted measure of the labour income share, as well as several adjusted measures. The figure highlights several key patterns. The panel on the left makes clear that, even in a developed economy such as the United States, the assumption made by the G3 measure is at odds with the microdata evidence. The G3 measure assumes that the relative wage of the self-employed is one, yet the evidence points to relative wages well above one. In 2016 for instance, the unadjusted labour share stood at 53.7 per cent. The G3 measure of the adjusted labour share was 57.3 per cent, whereas the microdata based measure was 59.0 per cent. This is consistent with the idea that in high income countries, self-employed workers tend to earn a higher relative wage than employees, and hence there is a self-employment premium. The magnitude of the premium is sizeable as is the effect it has on the adjusted labour share. In the present example, the additional effect of the self-employment premium represents roughly half of the G3 adjustment. The fact that in the United States the effect of accounting for relative wage differences is of the same order of magnitude as accounting for self-employment at all, highlights the importance of estimating relative wages in high income countries. Furthermore, the effect is also important from a dynamic perspective. The gap between the G3 and the ILO measure has decreased by roughly 20 per cent since 2005, suggesting that the labour share has fallen by more than is commonly reported. The effect is modest given the short horizon considered, but it can have a large impact when considering longer spans. The evidence suggests that currently used rules of thumb for developed countries will bias both the level and the evolution of the labour share. In fact, the time variation renders any fixed parameter assumption vulnerable to substantial errors, particularly over long horizons. Such spans are common in the empirical labour income share literature, for instance Karabarbounis and Neiman (2014) analyse the evolution of the labour share in several countries between 1975 and 2014.

The panel on the right of Figure 6 shows the different labour share measures for India. Clearly, assuming that the self-employed earn the same labour income than employees in income is not acceptable. In fact,

the estimated G3 measure lays above 100 per cent. But even following a rule of thumb suggested for developing economies, Treeck (2017), assuming a relative wage of two-thirds, the labour share would incorrectly exceed 100 per cent. Therefore, if the United States example shows the relevance of basing relative wages on microdata in high income countries, India's example shows that in developing countries the exercise is a necessity. The reason for the implausible results is straight-forward, the share of self-employment in India (and many other developing countries) is high. At the same time, microdata estimates suggest that there is a large self-employment penalty in India (in 2005 the wage of an own-account worker is estimated to be roughly a fifth of an employee's wage). The combination of large shares of self-employment together with large self-employment penalties renders rules of thumb almost unusable. First, there is a large risk of overestimating the labour share, as exemplified by above 100 per cent shares. Second, given the importance of self-employment in overall employment, the relative wage is the main determinant of the adjusted labour income share. Hence, the assumption behind the rule of thumb is in fact what drives the results – greatly limiting the value of such estimates. Third, both self-employment as well as the self-employment penalty can follow strong trends in the developing world. Therefore, the use of fixed rules of thumb are prone to unacceptable margins of error. At the same time, India's panel in Figure V makes a strong case in favour of adjusting for self-employment. In 2005 the adjusted share stood at 58 per cent, whereas the unadjusted share was 27.5 per cent. In a country where above 80 per cent of the work force was self-employed, the assumption that all those workers do not earn any labour income is not practical. This evidence, as well as the discussion in the methodological section concerning mixed income approaches, makes a forceful argument in favour of microdata based approaches.

Figure 6. The microdata based adjustment vs. rule of thumb (G3 and variants)

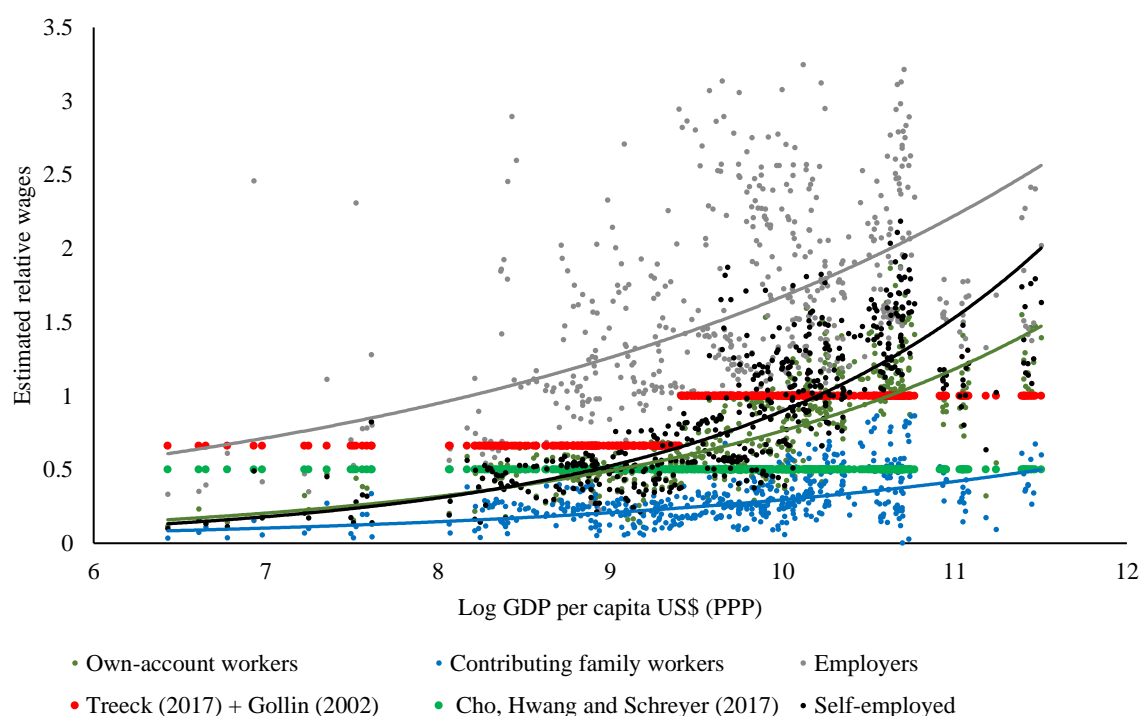


Note: The figure plots the unadjusted labour income share, the adjusted labour income share based on the current proposal (ILO\_LIS), and the 3<sup>rd</sup> adjustment proposed by Gollin (G3\_LIS). In the case of India, since it is a developing economy, the variant of G3 suggested by Treeck (2017), which proposes a fixed relative wage of the self-employed of two-thirds, (VT\_LIS) is also plotted.

The advantages of using microdata estimates of the relative wages of the self-employed, as opposed to rules of thumb, can be seen clearly when analysing the relationship of the relative wages with income by detailed status, and comparing them to an illustration of a rule of thumb (Figure 7). One key takeaway is the importance of not considering economic development as a binary status, developed vs developing. In fact, in any of the relative wages estimated there does not seem to be discontinuity, or step-like transition. Instead, the relative wage of the self-employed steadily increases with income (the relationship is reasonably well described by an exponential fit). A relative wage of one in developed economies understates the estimated labour income, and of 0.67 in developing countries overstates it (and so does the 0.5 rule). Naturally, rules of thumb that do not differentiate by level of economic development tend to miss the microdata estimates by even a larger margin. Furthermore, the complexity within self-employment is also shown to be of great importance. A relative wage of one in developed

economies is at least a good guess of the average microdata estimate for own-account workers. Nonetheless, employers and contributing family workers present a completely different relative wage than own-account workers. The former group tends to present higher relative wages than own-account workers. Unsurprisingly, contributing family workers are estimated to present the lowest relative wage. In fact, the fitted contributing family workers relative wage in high income countries is lower than the fitted value of employers in the poorest countries. Given the wide availability of detailed status in employment, it seems convenient to take into consideration the self-employment composition to adjust the labour income share. Overall, it can be concluded that both the complexity of self-employment as well as the continuous relationship between economic development and relative wages severely limit the usefulness of rules of thumb – which fail to capture average microdata estimates. Additionally, the variation around the fitted line can reflect idiosyncratic country characteristics that by construction rules will not be able to capture.

Figure 7. Estimated relative wages of the self-employed by detailed status in employment vs income: microdata estimates and rules of thumb



Note: The figure shows the estimated relative wage using microdata. The y-axis represents the relative wage (and is cut at 3.5 for visualization purposes) of the self-employed. The x-axis represents the log of GDP in PPP\$ per capita.

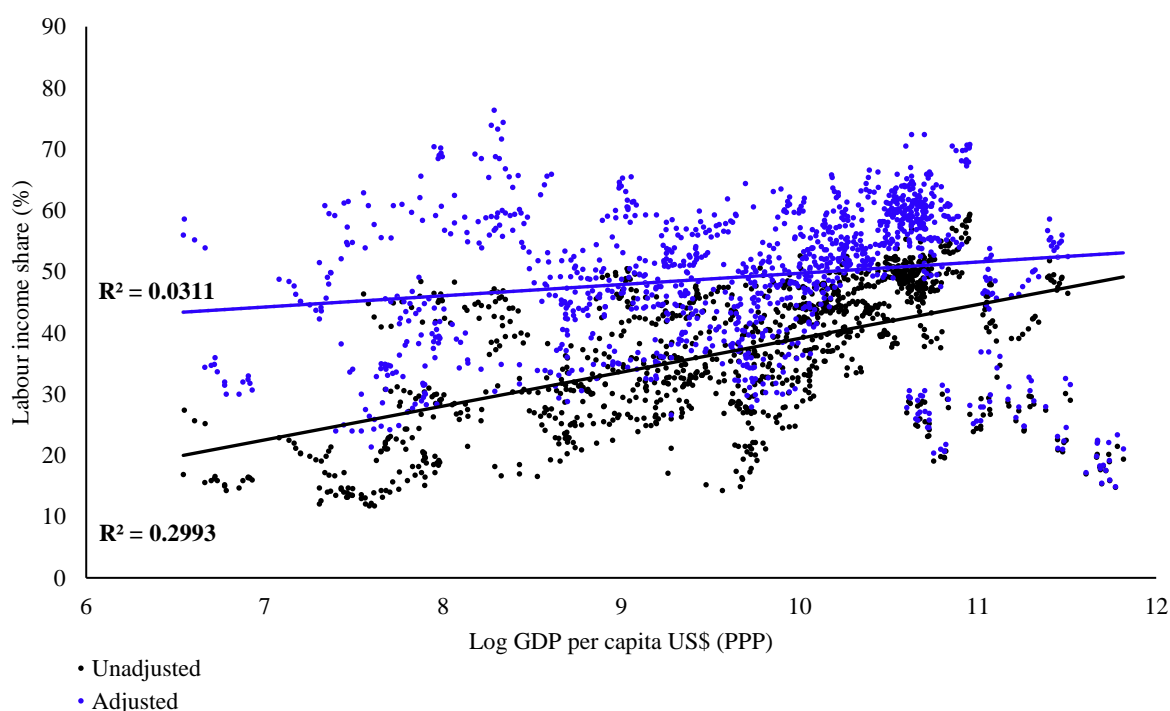
One of the reasons to adjust the labour income share for self-employment is to avoid an excessive correlation with income due to the fact that richer countries tend to have larger employee shares. Indeed, the labour share reflects important structural information of an economy, for instance: the aggregate technology function or the degree of monopoly power. Not adjusting for self-employment would present the risk of drawing the wrong conclusions on such structural parameters in the developing world. In this study, the limitations of rules of thumb, such as the one proposed by Gollin (2002), are shown. Nonetheless, in the same study the author claims that:

It has become widely accepted, in recent years, that labor shares are lower in poor countries than in rich countries. This has led to numerous ad hoc adjustments in growth models and trade models. This paper suggests that, for many analyses, it is reasonable to use models that give rise

to constant factor shares. [...] Estimates of factor shares that do not account for self-employment income will be seriously flawed, especially in poor countries.

The results of the present study point to only a very mild relationship between labour shares and income and certainly the findings support strongly Gollin's view concerning the need to adjust the labour income share, particularly in poorer countries, but also in developed economies. Figure 8 plots the adjusted and unadjusted labour shares vs log GDP per capita. The differences in the linear fit, if the adjusted measure is taken at face value, reflect the expected bias of unadjusted labour shares by level of income. The difference in the adjusted labour share is around ten percentage points between the richest and the poorest countries in the sample (in contrast to the 30 percentage points of the unadjusted measure). Furthermore, the overall role of income in determining the labour share is substantially weakened: whereas GDP per capita can explain 30 per cent of the variance in the unadjusted measure; it can only account for three per cent of the variance in adjusted labour income shares. The importance of the adjustment is not only apparent in the cross-section. The dynamic effect of the self-employment adjustment is, even in the short horizon of 13 years, quite significant. Figure 9 presents the trend of the unadjusted and adjusted shares. The trend estimate is based on the estimated yearly fixed effects of a GDP-weighted regression of 39 countries. The countries have been selected on the basis of having a complete time series of the unadjusted labour share, and at least one microdata based estimate of own-account workers. By construction the 2004 effect is set to 0. The results are clear, the adjustment significantly increases the downward trend of the labour income share. For these 39 countries, the adjusted measure presents a GDP weighted average decline of 3.4 per cent, whereas the unadjusted measure is just of 1.6 per cent. In particular, the decoupling between the two measures can be observed in the aftermath of the financial crisis (after 2010).

Figure 8. Unadjusted and adjusted labour shares vs GDP per capita



Note: Only countries with adjusted and unadjusted measures of the labour share are sub-sampled for this figure. The y-axis represents labour share measures. The x-axis represents the log of GDP in PPP\$ per capita.

Figure 9. Unadjusted and adjusted labour shares trends, 2004-2016, year fixed effects

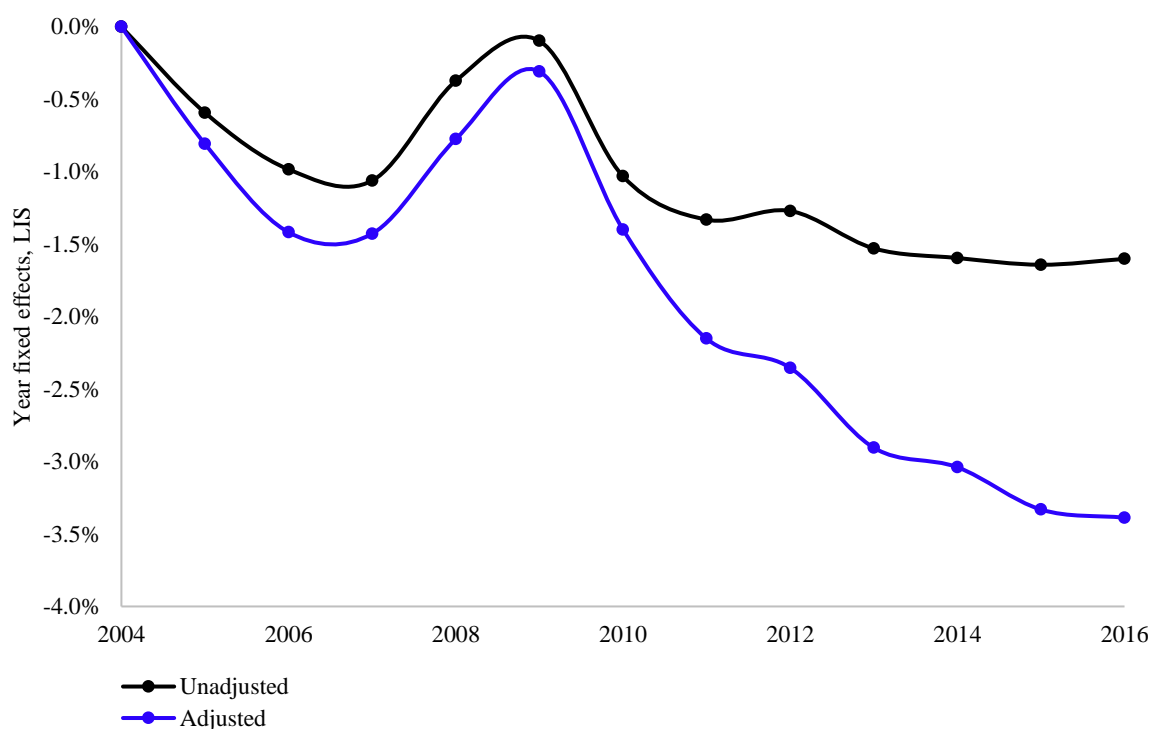


Table 2 describes the complete set of adjusted labour share estimates. The relative dispersion of the adjusted LIS is substantially lower than the unadjusted measure. The relative dispersion, the standard deviation divided by the mean, of the adjusted share is (for observations with a complete set of data) 17 per cent, while the unadjusted one is 30 per cent. Importantly, the average adjusted labour income share is substantially lower than the commonly cited two-thirds standard. For countries with all required data, the adjusted LIS is 53.1 per cent, whereas for countries with unadjusted labour income share data but with at least one of the variables necessary for the correction missing is 49.1 per cent. Finally, for countries without the raw data, the average adjusted share is 45.7 per cent. The smaller magnitudes in the last two groups are consistent with the positive relationship between income and data availability, and a negative one between labour income share and income.

Table 2. LIS: unadjusted and adjusted by method

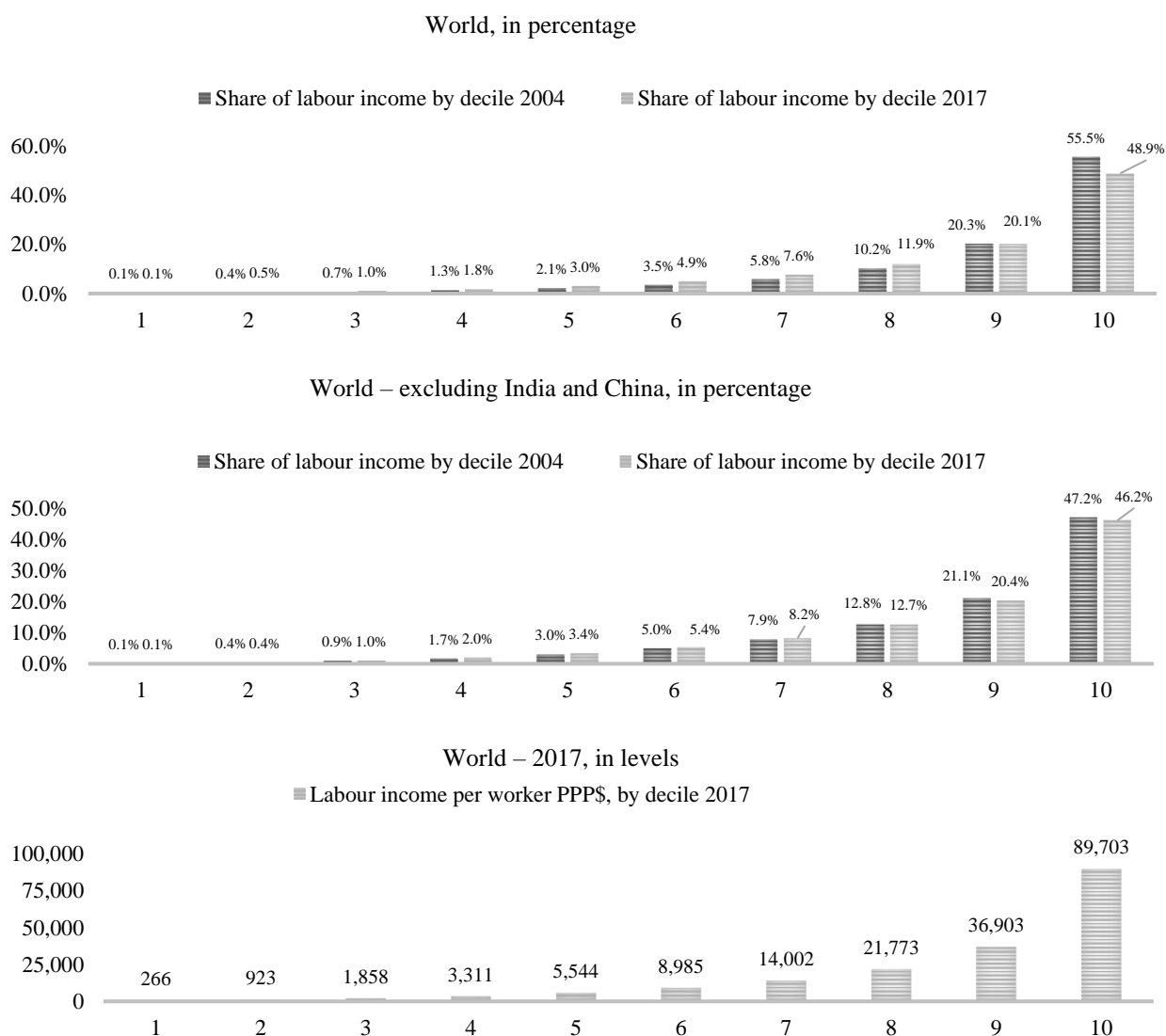
Measure	Observations	Mean	S.D.	P5	P95
Unadjusted LIS	1 212	37.2	11.1	20.7	51.5
Adjusted LIS (Microdata approach)	2 632	47.4	10.0	29.6	62.5
Imputed	1 420	45.7	8.2	30.8	59.2
Corrected	1 212	49.1	11.4	28.0	64.4
Complete set of data	329	53.1	8.8	35.9	65.9

Note: Imputed indicates that no data on the unadjusted labour income share was available. Corrected indicates that such data was used. Complete set of data indicates that all the necessary variables to compute the adjusted measure were available, and hence no imputation procedure took place. The variables include: unadjusted labour income share, detailed status in employment, and the microdata estimated relative wages of employers, own-account workers, and contributing family workers. The table excludes observations from China given the ad-hoc methodology applied to the country and explained in the appendix. The dispersion measures are respectively: standard deviation, 5<sup>th</sup> percentile and 95<sup>th</sup> percentile.

## 2. THE DISTRIBUTION OF LABOUR INCOME

Global labour income is very unequally distributed. In 2017, the top decile earned 48.9 per cent of all labour income, whereas the poorest ten per cent of the global workforce earned 0.1 per cent. Furthermore, the poorest 50 per cent of global workers earn just 6.4 per cent of the world's labour income. Nonetheless, as Figure 10 shows, global inequality has decreased in the last decade. In 2004 the richest decile accounted for 55.5 per cent of earnings, whereas the bottom 50 per cent earned 4.6 per cent of overall labour income. The key driver of the decline in labour income inequality is above average economic growth rates in certain developing countries, particularly India and China. Excluding India and China, the results point to a much slower reduction in labour income inequality: the poorest 50 per cent earned 6.9 per cent of labour income in 2017 compared to 6.1 per cent in 2004. Similarly the top ten per cent earned 46.2 per cent in 2017, a moderate decline from the 47.2 per cent in 2004. This does not indicate that in India or China inequality has decreased, indeed neither country registered a decline in inequality in the 2004-2017 span. Nonetheless, the two countries have experienced very high growth, which together with their initial low labour income level contributes to a global decrease in inequality. The last panel illustrates the degree of inequality in the distribution of labour income. A worker in the global top decile (a group of approximately three hundred million people) earned an average of roughly US\$89'000 (PPP) in labour income, whereas the global bottom decile earned US\$266 (PPP).

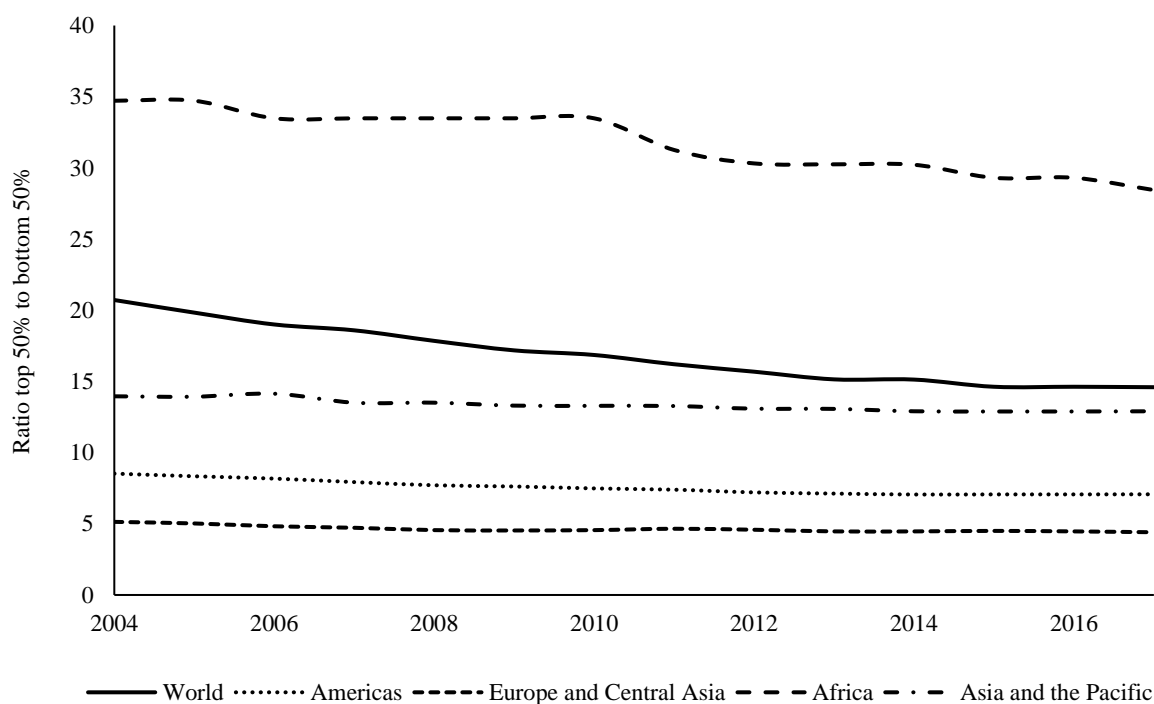
Figure 10. Labour income distribution per decile





A useful measure of inequality is to compute the ratio of the top 50 per cent of the distribution to the bottom 50 per cent. One can interpret this measure as the number of years that the poorest half needs to work on average to earn the same as the richest half earns in a year. Figure 11 plots the global magnitude as well as regional trends. The heterogeneity in inequality across regions is quite apparent – contrary to the previous section results on the labour income share. Worldwide, the poorest half of the employed population had to work around 14 years to earn the same that the richest half in 2017. In Africa the magnitude increases to 28, whereas in Europe it is roughly 4. The figure also shows that it is in Africa where inequality reduction has been more steep, nonetheless it has somewhat stalled since 2013. The global labour income inequality has mirrored this trend. In Asia and the Pacific and in Europe the stall started earlier.

Figure 11. Ratio of the labour income of the top 50 percent to the bottom 50 percent

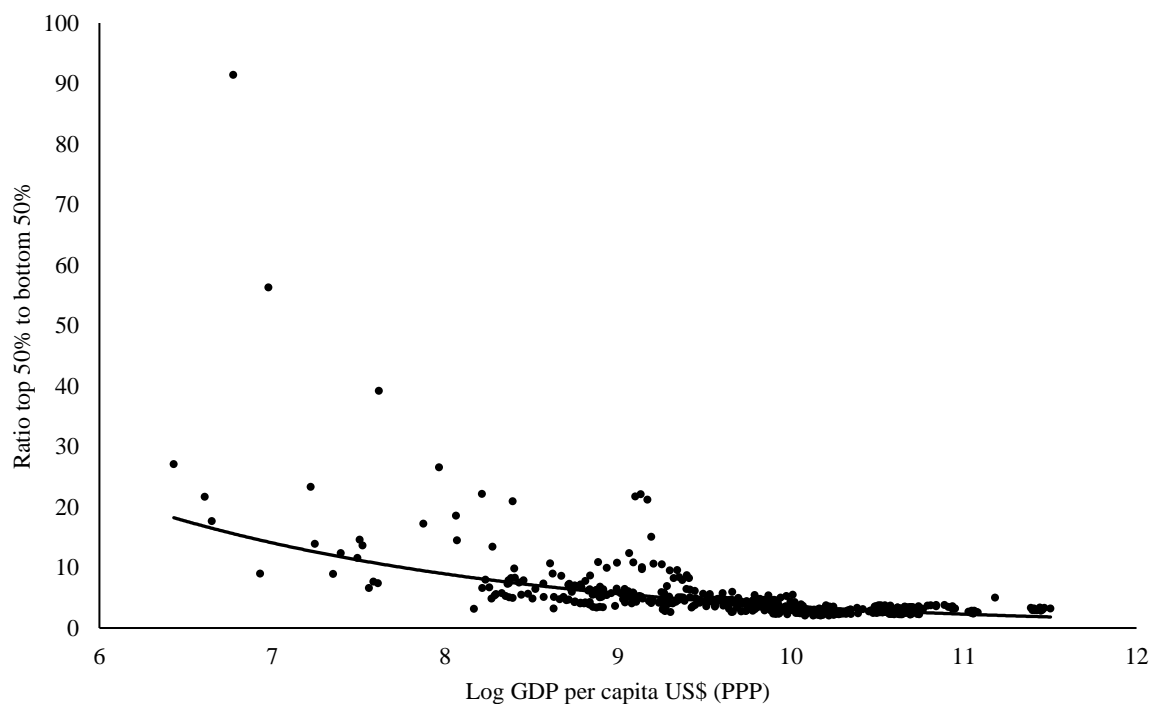


Given the highly unequal distribution of labour income worldwide, it is interesting to analyse the issue further. In particular, the large degree of heterogeneity in GDP per capita raises the following question: is labour income more unequal than we would expect given the country differences in income per capita? The answer is yes. Figure 12 presents the same metric than the previous figure (the ratio of the top 50 per cent labour income over the bottom 50 per cent). The actual microdata estimated distributions of each country, instead of the regional averages, are plotted against the (log) GDP per capita of the observations. The results show a strong negative association between inequality and income per capita. Poor countries present much more unequal labour income distributions,<sup>53</sup> exacerbating the level of inequality that would result if no correlation was apparent. Therefore, labour income is unequally distributed globally both due to differences in average per worker labour income and more unequal distribution in countries with lower average income. The countries in the figure with the most unequal distributions (a ratio above 25) are Congo DR, Côte d'Ivoire, Liberia, Niger and Uganda. The high inequality of labour income in developing countries derives both from a top end of the distribution with very large incomes (say the top 10 per cent vs. the following 40 per cent) and a large share of the workforce (broadly speaking the bottom 50 per cent of the distribution) with extremely low labour

<sup>53</sup> It is also worth highlighting that poor countries will also tend to have smaller labour income shares, further increasing overall inequality.

income. The striking difference in the inequality of labour income distribution in developing countries highlights the advantage of complementing overall inequality measures with labour inequality measures. Not only because the latter can differ from the former, but because due to data and methodological constraints, focusing on labour income allows to estimate inequality for countries where no overall distributional data is available.

Figure 12. Ratio of the labour income of the top 50 per cent to the bottom 50 per cent vs (log) GDP per capita US\$ (PPP), Microdata estimated distributions

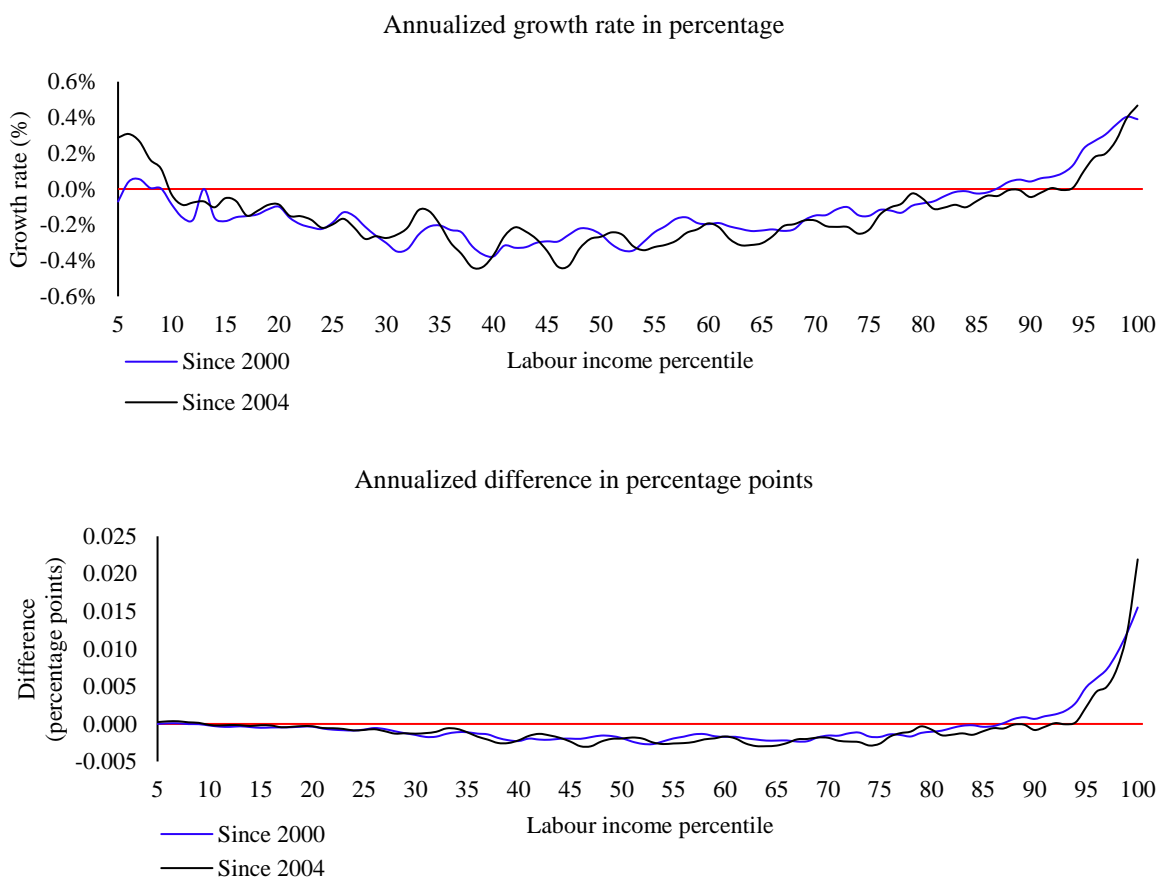


Note: The fitted line follows an exponential function

The estimates of labour income distribution at the country level allow to focus the attention to labour income inequality, and not just overall inequality. Distributional data in high income countries, with a focus on inequality, has been of particular interest recently. A developed country which has been recurrently studied is the United States. For instance, Piketty, Saez and Zucman (2018) find that in the last decades, the distribution of income has become more unequal. Strikingly, the authors find that the growth rate of income increased monotonically with income. The present study has a more limited time span, only starting in 2004. Nonetheless, the finding of a substantial decline in the labour income share already points to increasing inequality – since capital income tends to disproportionately benefit the rich. But the focus is on whether solely considering the labour income distribution adds additional information concerning inequality. Figure 13 presents the change in the distribution of labour income in the United States. Since 2004, the relationship between the growth rate and percentile for labour income is u-shaped, albeit the u is non-symmetrical and favours the higher end of the distribution. Both the lowest and highest labour income deciles saw their shares increase, at the expense of the rest of the distribution. Thus, roughly 80 per cent of the workforce, which would broadly encompass the middle class, has seen their share in overall labour income decline. Expanding the horizon to 2000-2016, reverts the increases of the lowest end of the distribution. Nonetheless, the fact that the middle classes are affected with the largest relative declines still holds. Focusing only on relative growth rates can obscure the main drivers of the change in the distribution. To highlight these, the figure also presents the same data, but plots the difference in percentage points. In this case, the resulting figure follows a “hockey-stick” pattern, with large increases in the shares at the top end of the distribution with losses for the rest,

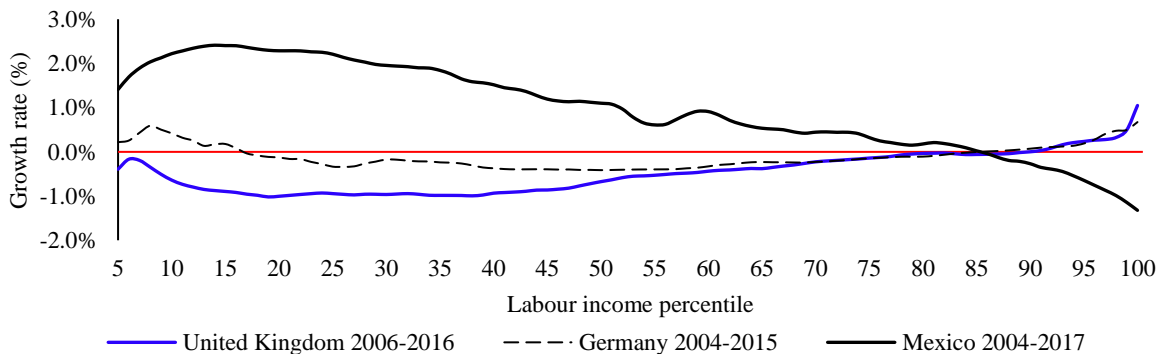
particularly for the middle of the distribution. The driving force behind this is outside the scope of the present study. The pattern is being raised as an example of new interesting trends and facts that can emerge when focusing solely on labour income. Figure 14 presents the same exercise (only considering growth rates) for a group of three other countries based on their economic size, the diversity of changes in distribution, and a rich availability of microdata. The United Kingdom presents a similar pattern to the one observed in the United States, but with the largest declines being observed between the 10<sup>th</sup> and 40<sup>th</sup> percentile respectively, and the lowest end of the distribution clearly declining in shares (but less than most of the rest). Germany, in contrast, presents a u-shape markedly clearer than the United States. Mexico shows a clear reduction in labour income inequality (in contrast with the strong decline that it experiences in the labour income share).

Figure 13. Labour income distribution change by percentile, United States 2004-2016



Note: Given the volatility the data is presented as a moving average with a period of five to smooth results.

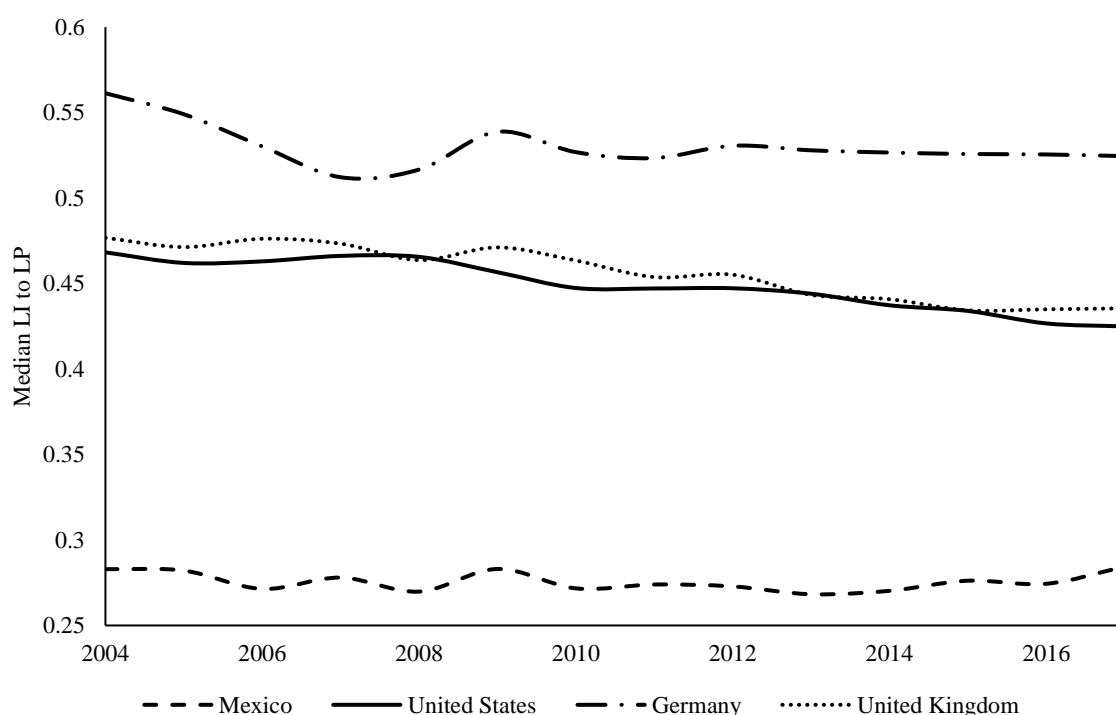
Figure 14. Annualized growth rate by percentile: Germany, Mexico and United Kingdom



Note: Given the volatility the data is presented as a moving average with a period of five to smooth results.

The results in Mexico show clearly the importance of analysing jointly the behaviour labour income share and the labour income distribution. For the aforementioned four economies, Figure 15 plots the ratio of the median labour income to labour productivity (GDP in US\$ (PPP) per worker). This measure has a high synthetic value since it allows to examine both the effects of changes in the labour income share and the distribution of the labour income. If the income accruing to capital was 0, and all workers earned the same labour income, the ratio would be equal to 100 per cent. Nonetheless, rather than the level it is more interesting to focus on its evolution. The United States and the United Kingdom present a clear decline in the ratio, due to the compounding effects of declining labour shares and shrinking of the share of labour income accruing to the median worker. Germany presents a smaller decline, largely thanks to the slower decline of the overall labour share. Finally, in Mexico the decreasing inequality in the distribution of labour income coupled with the decline in the labour share produce a more or less stagnant ratio.

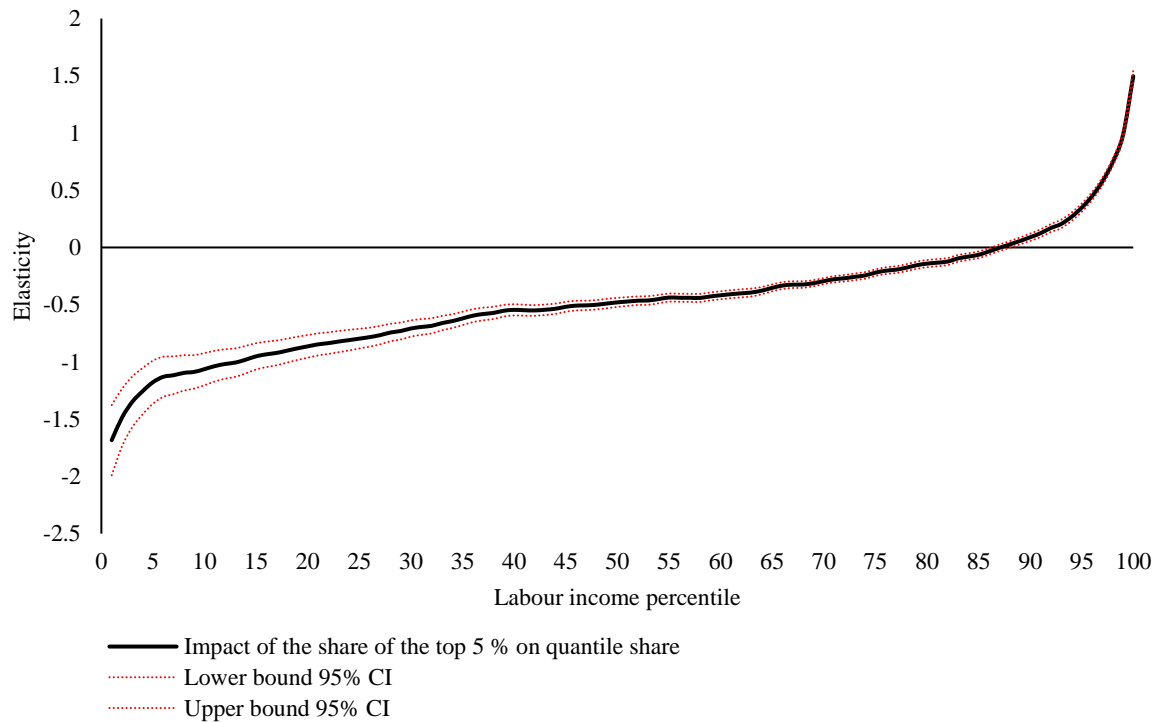
Figure 15. Median labour income divided by labour productivity, 2004-2017



Given the variety of patterns in the labour income distribution, it is worth exploring if some common tendencies arise. For instance, it is interesting to assess whether increases in the top are associated with losses for all the rest, or the zero sum changes disproportionately affect particular segments of the distribution. To do so a fixed effect regression is set up in which the independent variable (in logs) is either the median share of labour income (approximated as the average of the 50<sup>th</sup> and 51<sup>st</sup> percentiles) or the share of labour income of the top five percent. The dependent variable is the (log) share of labour income of each percentile. Hence, by running a within effects regression for each percentile, the average within country elasticity of the share of income by percentile with respect to the independent variable is obtained. The selected sample only includes the countries for which at least two microdata estimated labour income distributions are available. In Figure 16 it can be clearly seen that increases in the shares of the top five per cent are associated with declines for much of the rest of the distribution. In particular all the percentiles of labour income between the first and the 86<sup>th</sup> percentiles are found to experience decreases in their shares. The magnitude is statistically and economically significant. An increase of 1 per cent in the top 5 per cent share is associated with a decline of 0.4 per cent for the 50<sup>th</sup> percentile and

1.6 per cent for the first one.<sup>54</sup> In a similar way, Figure 17 shows that increases in the median share of labour income are positively associated with increases in almost all the distribution, excluding approximately the 87<sup>th</sup> percentile and above. The positive association is particularly strong towards the end of the distribution. Hence, the association shows that increases in the middle of the distribution have a disproportionately strong lifting effect for the poorest workers.

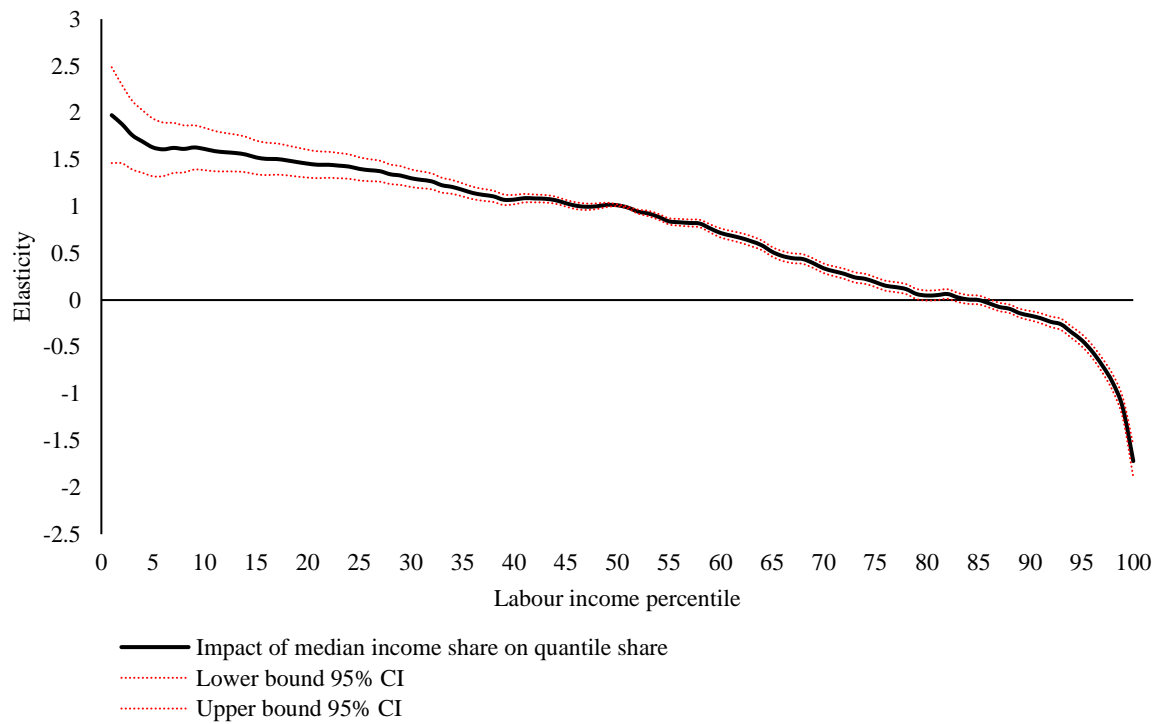
Figure 16. Within country elasticity of each percentile share of labour income to the labour income share of the top five per cent



Note: The figure plots the result of 100 regressions between the dependent variable, log share of labour income per percentile, and the independent variable, log share of labour income of the top five per cent, as well as country fixed effect controls.

<sup>54</sup> This is not a mechanic result, even if the cumulative impact of the changes must be consistent with 0 net change, increases in the top 5 per cent share could very well be correlated to increases in the middle or low end of the distribution and decreases elsewhere.

Figure 17. Within country elasticity of each percentile share of labour income to the median share of labour income



Note: The figure plots the result of 100 regressions between the dependent variable, log share of labour income per percentile, and the independent variable, log share of the median share of labour income (approximated as the average of the 50<sup>th</sup> and 51<sup>st</sup> percentiles), as well as country fixed effect controls.

## 4. CONCLUSION

---

The labour income share is a macroeconomic variable and an inequality measure that has raised substantial social and academic interest. The measure is included as an indicator to measure inequality in the United Nations Sustainable Development Goals framework. At the same time, the literature has studied extensively the stability, long-run trends and cyclical properties of the labour income share. Similarly, the interest in detailed distributions of income, as well as other key metrics of inequality such as poverty estimation, has risen steadily both in developing economies and in advanced economies, particularly since the aftermath of the 2008 financial crisis. Nonetheless, the estimation of inequality metrics has encountered substantial data challenges.

The main challenge in measuring the labour share is how to estimate the labour income of the self-employed. Whereas the share of income accruing to employees can be readily obtained via the national accounts item, compensation of employees, the labour income of the self-employed cannot. The difficulties are conceptual and practical: the extent to which the income of self-employed workers accrues to labour or capital is a notion that cannot be directly produced from the data. Similarly, the overall income of the self-employed as identified in national accounts, mixed income, presents both a higher methodological heterogeneity across countries and a much higher degree of uncertainty than compensation of employees. To solve this, the literature has often cited the use of observation level data, household survey microdata, to impute the labour income of the self-employed, taking as a basis the compensation of employees. This type of exercise has been seldom performed, in favour of rule of thumb approaches, due to scarce access to the necessary household surveys. In parallel, the estimation of detailed income distribution – beyond aggregate measures such as the labour share – has captured the attention of the empirical inequality literature. One pervasive problem found is that household surveys do not capture household income well, due to, for instance, the systematic underestimation of income for high earners that causes under coverage with respect to national accounts and a biased income distribution. Thus, whereas in the labour income share estimation household surveys are thought of as an improvement of the measurement process, the estimation of the income distribution literature is looking to administrative or national accounts sources as substitutes or correction factors for household surveys.

This study proposes to bridge the gap between the two strands of literature by treating the different types of income, capital and labour income, separately. Taking advantage of the ILO Harmonized Microdata collection, the labour income of the self-employed is estimated to produce a labour income share that does not rely on rules of thumb, based on household surveys from 95 countries. This allows to split the economy in a labour income component and a capital one. Afterwards, the question of whether the distribution of labour income can be reasonably estimated using household surveys is examined, even if they are not accurate enough to measure the overall income distribution. A country case study is carried out for Belgium, benchmarking the distribution of employee compensation from a household survey with the one from an establishment survey (that randomly draws earnings data from the social security register) that shows that the household survey closely approximates the benchmark. In a similar manner, existing evidence points to a high degree of consistency between labour survey data and national accounts in terms of coverage rates (the ratio of aggregate labour income by type of source). Moreover, the gaps in coverage rates (the mismatch in the total sum of labour income between sources) do not appear to systematically bias the labour income distribution in terms of over- or under-estimating inequality, in contrast to overall income. Thus it seems that the problems of estimating income distribution from household surveys are a consequence of bias in the estimation of capital income, and not to a substantial degree, labour income. Nonetheless, a systematic cross check of household surveys against administrative registers and establishment surveys would be of great interest. Given the data and metadata requirements of this exercise, it is left for future research.

The resulting dataset, which allows to identify by country the labour income, its distribution, and the capital income, has been made public on ILOSTAT. The main findings can be summarised as follows: (i) the global labour income share is declining and countercyclical, similar patterns arise in the European Union and the United States; (ii) the effects of self-employment on the labour income share are highly heterogeneous – highlighting the limitations of widely used rules of thumb; (iii) labour income inequality decreases strongly with national income level, hence average cross-country differences are greatly exacerbated; (iv) within countries, relative increases of labour income at the upper end of the distribution are associated, on average, with relative losses for the rest.



## 5. REFERENCES

---

- Alvaredo, F.; Atkinson, A.; Chancel, L.; Piketty, T.; Saez, E.; Zucman, G. 2016. "Distributional national accounts (DINA) Guidelines: Concepts and Methods used in WID.world," Working Papers 2016/02, World Inequality Lab.
- Anand, S.; Segal, P. 2008. "What Do We Know about Global Income Inequality?" *Journal of Economic Literature*, 46 (2008), 57–94.
- Appleton, J. 2011. "Revised Methodology for Unit Wage Costs and Unit Labour Costs – Explanation and Impact". ONS, pages one to ten.
- Blanchet, T., Flores I., Morgan M. 2018. "The Weight of the Rich: Improving Surveys Using Tax Data", Working Papers 2018/12, World Inequality Lab.
- Blanchet, T.; Fournier, J.; Piketty, T. 2017. "Generalized Pareto Curves: Theory and Applications," Working Papers 2017/03, World Inequality Lab.
- Bridgman, B. 2014. "Is Labor's Loss Capital's Gain? Gross versus Net Labor Shares," BEA Working Papers 0114, Bureau of Economic Analysis.
- Bentolila, S.; Gilles, S.P. 2003. "Explaining Movements in the Labor Share," *The B.E. Journal of Macroeconomics*, De Gruyter, vol. 3(1), pages 1-33, October.
- Cabral, A.C.; Norman, G. 2018. "Estimating Self-Employment Income-Gaps from Register and Survey Data: Evidence for New Zealand," Working Paper Series 7625, Victoria University of Wellington, Chair in Public Finance.
- Crespi, G. 2004. "Imputation, estimation and prediction using the Key Indicators of the Labour Market (KILM) data set". Employment Strategy Paper 2004/16, (Geneva, ILO)
- Chen, S.; Ravallion, M. 2010. "The Developing World Is Poorer Than We Thought, but No Less Successful in the Fight against Poverty," *Quarterly Journal of Economics*, 125 (2010), 1577–1625
- Cho, T.; Hwang, S.; Schreyer, P. 2017. "Has the Labour Share Declined?: It Depends," OECD Statistics Working Papers 2017/1, OECD Publishing.
- Deaton, A. 2005. "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)," *Review of Economics and Statistics*, 87 (2005), 1–19.
- Deninger, K; Squire, L. 1996. "A New Data Set Measuring Income Inequality," *The World Bank Economic Review* Vol. 10, No. 3 (Sep., 1996), pp. 565-591 (27 pages)
- De Stefano, V. 2016. "The rise of the 'just-in-time workforce': on-demand work, crowdwork and labour protection in the 'gig-economy'," ILO Working Papers 994899823402676, International Labour Organization.
- Feenstra R.C; Inklaar, R.; Timmer, M.P. 2015. "The Next Generation of the Penn World Table," *American Economic Review*, American Economic Association, vol. 105(10), pages 3150-3182, October.
- Fesseau, M.; Wolff, F.; Mattonetti, M.L. 2013. "A Cross-country Comparison of Household Income, Consumption and Wealth between Micro Sources and national accounts Aggregates," OECD Statistics Working Papers 2013/3, OECD Publishing.
- Galí, J.; Gertler, M. 1999. "Inflation dynamics: A structural econometric analysis," *Journal of Monetary Economics*, Elsevier, vol. 44(2), pages 195-222, October.

- Gammarano, R. 2018. "Paid employment vs vulnerable employment," ILOSTAT Spotlight on work statistics, International Labour Organization.
- Gollin, D. 2002. "Getting Income Shares Right," *Journal of Political Economy*, University of Chicago Press, vol. 110(2), pages 458-474, April.
- Heckman, J. 1979. "Sample Selection Bias as a Specification Error," *Econometrica*, Econometric Society, vol. 47(1), pages 153-161, January.
- ILO, 1993. "Revision of the International Classification of Status in Employment," ILO.
- ILO, 2017. "ILO labour force estimates and projections: 1990-2030, Methodological description", ILO.
- ILO, 2018a. "ILO Modelled Estimates, November 2018, ILOSTAT", Research Department and Statistics Department, ILO.
- ILO, 2018b. "Women and men in the informal economy: A statistical picture," ILO.
- ILO, 2018c. "ILOSTAT microdata processing quick guide," Statistics Department, ILO.
- ILO, 2019. "Work for a brighter future – Global Commission on the Future of Work International Labour Office," ILO.
- IMF, 2017. "Understanding the Downward Trend in Labour Income Shares," Chapter three in *World Economic Outlook*, IMF.
- Johnson, T. 2003. "The underground economy and Australia's GDP," OECD.
- Karabarbounis, L.; Neiman, B. 2013. "The Global Decline of the Labor Share," *Quarterly Journal of Economics*, Vol. 129(1), pp.61-103.
- LIS, 2019. "Luxembourg Income Study (LIS) Database," <http://www.lisdatacenter.org>, LIS.
- Kapsos, S., Bourmpoula, E. 2013. "Employment and economic class in the developing world," Research Paper No. 6, ILO.
- Krueger, A.B. 1999. "Measuring Labor's Share," *American Economic Review*, American Economic Association, vol. 89(2), pages 45-51, May.
- Milanovic, B. 2002. "True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone," *Economic Journal*, 112(476): 51–92.
- Milanovic, B. 2005. "Worlds Apart: Measuring International and Global Inequality," Princeton, NJ: Princeton University Press.
- Milanovic, B. 2015. "Global inequality of opportunity: how much of our income is determined by where we live," *Review of Economics and Statistics*.
- ONS, 2018. "Estimating the impact of the self-employed in the labour share", ONS.
- Piketty, T. 2014. "Capital in the Twenty-first Century," Harvard University Press.
- Piketty, T.; Saez, E.; Zucman, G. 2018. "Distributional National accounts: Methods and Estimates for the United States," *The Quarterly Journal of Economics*, Volume 133, Issue 2, 1 May, Pages 553–609.
- Piketty, T.; Yang, L.; Zucman, G. 2018. "Capital Accumulation, Private Property and Rising Inequality in China, 1978-2015," HKUST IEMS Working Paper Series 2018-54, HKUST Institute for Emerging Market Studies, revised Mar 2018.
- Pinkovskiy, M.; Sala-i-Martin, X. 2016. "Lights, Camera ... Income! Illuminating the National Accounts-Household Surveys Debate," *The Quarterly Journal of Economics*, Volume 131, Issue 2, May 2016, Pages 579–631

Pionnier, P.A.; Guidetti, E. 2015. "Comparing profit shares in value-added in four OECD countries: Towards more harmonised national accounts," OECD Statistics Working Papers 2015/3, OECD Publishing.

Rothbaum, J.L. 2015. "Comparing Income Aggregates: How do the CPS and ACS Match the National Income and Product Accounts, 2007-2012," SEHSD Working Paper 2015-01.

Sala-i-Martin, X. 2006 "The World Distribution of Income: Falling Poverty and ... Convergence, Period," *The Quarterly Journal of Economics*, Volume 121, Issue 2, May 2006, Pages 351–397.

Solt, F. 2019. "Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database." SWIID Version 8.0, February 2019.

Trapp, K. 2015. "Measuring the labour income share of developing countries: Learning from social accounting matrices," UNU WIDER Working Paper 2015/041, 1-24

Van Treeck, K. 2017. "The Labour Income Share in Developing Countries: A Review and Analysis of International Panel Data," 5th Conference of the Regulating for Decent Work Network at the International Labour Office.

UNSD, EC, IMF, OECD, WB, 2009. "System of National Accounts, 2008," United Nations.

WB, 2013. "World Development Report 2013," World Bank.

Young, A. 1995. "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience," *The Quarterly Journal of Economics*, Oxford University Press, vol. 110(3), pages 641-680.

## 6. APPENDIX

---

### 1. ADAPTATION OF THE METHODOLOGY: THE CASE OF CHINA

The methodology described in this study is applied to all the countries in the sample except China.<sup>55</sup> This is because the country presents very scarce internationally comparable employment data. At the same time, given its size both in terms of employment and GDP the country is a key driver of global results. The data limitations are the following. Internationally comparable microdata are not publicly available. Comparable status in employment aggregate data are not available either. The nationally reported self-employment is very low (considering the GDP per capita of the country), and it is quite volatile. Its evolution is at odds with the international evidence in high growth countries. Furthermore, national accounts data reported to UNSD do not include mixed income, whereas the compensation of employees is quite high given the GDP per capita of the country. Moreover, according to Piketty, Yang and Zucman (2017) the compensation of employees includes all income (labour related earnings + operating surplus) of farmers. International evidence shows that at the GDP per capita level of the country, one should expect a substantial share of self-employment to be agricultural. Therefore it seems likely that a sizeable amount of labour income (and capital income) of the self-employed is already added to compensation of employees.

Given all these considerations, no adjustment is done for self-employment. Hence, the estimate of the adjusted labour income share is equal to the unadjusted share. This is clearly a blunt assumption, and one that could be overcome with access to the microdata of the country's labour force survey or at least with internationally comparable data of status in employment. Given the importance of the role of the employment distribution by status in imputing missing observations for the quantiles of labour income, an ad hoc procedure to estimate the quantiles of labour income for China is also necessary. This is due to the huge uncertainty associated with the variable of status in employment. The ad hoc procedure is based on running the same regression setting used for all the countries, but instead of using the share of employees, the GINI index of market income from Solt (2019) is used.

---

<sup>55</sup> For Vanuatu the direct correction of the unadjusted labour income share is not used, given that it results in a very high value (above 80 per cent). Instead, the adjusted share is estimated in the regression step – taking as input variable the unadjusted labour income share.

## 2. SELF-EMPLOYMENT DIFFERENCES BETWEEN NATIONAL ACCOUNTS AND ICSE

The main difference between the status in employment classification from ICLS or national accounts concerns a certain group of workers which according to ICLS standards are self-employed. The SNA 2008 UNSD manual Paragraph 7.28 reads: *“The definitions in the SNA are consistent with resolutions of the International Conference of Labour Statisticians (ICLS) concerning the definitions of the economically active population. For the SNA, though, the main objective is to clarify the nature of the employment relationship in order to fix the boundary between compensation of employees and other kinds of receipts. Some persons who in labour statistics may be included with the self-employed, in particular some owners of quasi-corporations and owner-managers of corporations, are treated in the SNA as employees”*. Furthermore in paragraph 19.24 the point is made more explicitly: *“Managers of corporations (or quasi-corporations) are treated in the SNA as employees but the ILO classification regards them as self-employed”*. And in 19.28: *“In ILO statistics, self-employed persons include those working in enterprises that are legally unincorporated even if there is sufficient information available for them to be treated as quasi-corporations in the SNA. In the SNA the remuneration of these people is included in compensation of employees rather than in mixed income”*.

Indeed the two standards differ, the 15th International Conference of Labour Statisticians (ICLS) defines self-employment in paragraph 7 as: *“Self-employment jobs are those jobs where the remuneration is directly dependent upon the profits (or the potential for profits) derived from the goods and services produced (where own consumption is considered to be part of profits). The incumbents make the operational decisions affecting the enterprise, or delegate such decisions while retaining responsibility for the welfare of the enterprise. (In this context “enterprise” includes one-person operations)”*. And even more specifically, in paragraph 14.a it is *“Owner-managers of incorporated enterprises are workers who hold a job in an incorporated enterprise, in which they:(a) alone, or together with other members of their families or one or a few partners, hold controlling ownership of the enterprise; and (b) have the authority to act on its behalf as regards contracts with other organizations and the hiring and dismissal of persons in “paid employment” with the same organization, subject only to national legislation regulating such matters and the rules established by the elected or appointed board of the organization. Different users of labour market, economic and social statistics may have different views on whether these workers are best classified as in “paid employment”(cf. paragraph 6) or as in “self-employment” (cf. paragraph 7), because these workers receive part of their remuneration in a way similar to persons in ‘paid employment’ while their authority in and responsibility for the enterprise corresponds more to persons in “self-employment”, and in particular to employers’. (Note, for example, that to classify them as “employees” will be consistent with their classification in the “System of National Accounts”, while they may be best classified as “employers” or “own-account workers” for labour market analysis.) Countries should, therefore, according to the needs of users of their statistics and their data collection possibilities, endeavour to identify this group separately. This will also facilitate international comparisons.”*

Therefore, those owners of incorporated enterprises (or of quasi corporations) whose remuneration is directly dependent upon the profits, make operational decisions affecting the enterprise and who hold a job in such enterprise will generally be considered self-employed according to ICLS. Nonetheless, in the SNA framework, the wages that they receive from their job in the enterprise will be included in the SNA compensation of employees item (D1), and the rest of the income that they receive from the enterprise will be considered as capital income. This circumstance cannot be taken as an economically meaningful division between labour and capital income, rather it is one that relates to administrative practices as well as taxation rules (for instance lower taxation on either capital or labour income would presumably strongly affect the relative shares of each type of income). If the data on the wages of these

workers would be separately available, one appropriate solution would be to subtract their remuneration from the SNA compensation of employees item (D1). However, this identification is not possible, since national accounts microdata are not readily available. Unfortunately, the second potential strategy of identifying the income of such workers in the ICLS based data is also not possible, since in general there are no identification data for these particular workers and furthermore no wage data for any type of self-employed are generally collected. Given these two limitations, this study is based on the assumption that the SNA and ICLS concepts are compatible to a reasonable degree, and ignores the actual mismatches. Inevitably, this leads to a certain amount of double counting. Addressing this problem is left for future research, based on having the access to the appropriate detailed national accounts data. Notwithstanding the above discussion, it must be highlighted that this potential mismatch does not appear to be a major determinant for the share of self-employed in total employment. One would expect that the ICLS-based self-employment, which includes owners with a job in their own corporation or quasi-corporation, to be upwardly biased, particularly in high income countries. Comparing the self-employment rate between labour force survey estimates and data from national accounts for 32 mostly high income countries<sup>56</sup> reported by Eurostat, there is no clear evidence of systematic bias in either direction, and substantial dispersion across countries.

---

<sup>56</sup> Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom.

Table 3 - Explanatory variables used in the regression to estimate the wages of employees.

Order	ISIC	ISCO	Urban	Education	Age	Gender	Hours
1	4-2D	08-2D	X	X	X	X	X
2	4-2D	88-2D	X	X	X	X	X
3	4-2D	08-2D		X	X	X	X
4	4-2D	08-1D	X	X	X	X	X
5	4-2D	Skill	X	X	X	X	X
6	4-1D	08-1D	X	X	X	X	X
7	4-1D		X	X	X	X	X
8	4-2D	08-2D	X		X	X	X
9	3.1-2D	88-2D	X	X	X	X	X
10	3.1-2D	88-1D	X	X	X	X	X
11	3.1-2D	88-2D		X	X	X	X
12	3.1-2D	Aggregate	X	X	X	X	X
13	3.1-1D	08-1D	X	X	X	X	X
14	3.1-1D	88-1D	X	X	X	X	X
15	Aggregate	Aggregate	X	X	X	X	X
16	Aggregate	Aggregate		X	X	X	X
17	Aggregate			X	X	X	X
18	Aggregate				X	X	X
19				X	X	X	X

Note: The order indicates the priority of the particular regression model, models with larger order numbers are only used if it is not possible to apply a model with a lower order number. The ISIC and ISCO variables indicate economic activity and occupation (ISIC: International Standard Industrial Classification of All Economic Activities and International; ISCO: International Standard Classification of Occupations). Several versions of the classification and level of detail are available in the underlying data sets. The two-digit (2D) or one-digit (1D) levels are used when available, and if not aggregate classifications are used instead. The hours of worked variable refers to either actual hours of work or usual hours of work (the former is chosen if both are available as it is the most common measure in the underlying data). The educational level is discretized into five categories: less than basic, basic, intermediate, advanced and level not stated. The urban variable indicates whether the worker resides in an urban or rural area. Finally, the variable age is the natural age in years (nonetheless due to sample size the ages 15-20 and 65+ are assigned the same two categories).

Table 4 - List of microdata sets used in the model

Reference Area	Survey	Reference Area	Survey
Argentina	EPH Encuesta Permanente de Hogares (Urban)	Liberia	LFS Labour force survey
Armenia	HLFS Household Labour Force Survey	Sri Lanka	LFS Sri Lanka Labour Force Survey
Austria	EUSILC Statistics on Income and Living Conditions	Lithuania	EUSILC Statistics on Income and Living Conditions
Burundi	ECVM Enquête sur les conditions de vie des ménages	Luxembourg	EUSILC Statistics on Income and Living Conditions
Belgium	EUSILC Statistics on Income and Living Conditions	Latvia	EUSILC Statistics on Income and Living Conditions
Benin	EMICOV Enquête Modulaire et Intégrée sur les Conditions...	Madagascar	ENESI Enquête Nationale sur l'Emploi et le Secteur Informel
Bangladesh	LFS Labour Force Survey	Maldives	HIES Household Income and Expenditure Survey
Bulgaria	EUSILC Statistics on Income and Living Conditions	Mexico	ENE Encuesta Nacional de Empleo
Bosnia and Herzegovina	LFS Labour Force Survey	Mexico	ENOE Encuesta Nacional de Ocupación y Empleo
Bolivia	ECE Encuesta Continua de Empleo	Mali	EMOP Enquête Modulaire et Permanente auprès des Ménages
Bolivia	EH Encuesta de Hogares	Malta	EUSILC Statistics on Income and Living Conditions
Brazil	PNAD Pesquisa Nacional por Amostra de Domicílios	Myanmar	LFS Labour force survey
Brazil	PNADC Pesquisa Nacional por Amostra de Domicílios Contínua	Mauritius	CMPHS Continuous Multi-Purpose Household Survey
Brunei Darussalam	LFS Labour force survey	Malawi	LFS Labour Force Survey
Canada	LFS Labour force Survey	Namibia	LFS Labour force survey
Switzerland	EUSILC Statistics on Income and Living Conditions	Niger	ECVM Enquête Nationale sur les Conditions de Vie de Ménages
Chile	CASEN Encuesta de Caracterización Socioeconómica Nacional	Nigeria	GHS General Household Survey
Côte d'Ivoire	ENSE Enquête nationale sur la situation de l'emploi	Nicaragua	ECH Encuesta Continua de Hogares
Cameroon	ECAM Enquête Camerounaise auprès des Ménages	Netherlands	EUSILC Statistics on Income and Living Conditions
Congo, Democratic ...	ENESI Enquête 1 2 3 sur l'Emploi et le Secteur Informel	Norway	EUSILC Statistics on Income and Living Conditions
Colombia	GEIH Gran Encuesta Integrada de Hogares	Nepal	LFS Nepal Labour Force Survey
Cape Verde	IMO Inquérito Multi Objectivo Contínuo Estatísticas ...	Pakistan	LFS Labour Force Survey
Costa Rica	ENAHO Encuesta Nacional de Hogares	Panama	EML Encuesta de Mercado Laboral
Cyprus	EUSILC Statistics on Income and Living Conditions	Peru	ENAHO Encuesta Nacional de Hogares
Czech Republic	EUSILC Statistics on Income and Living Conditions	Philippines	LFS Labour Force Survey
Germany	German Socio-Economic Panel (GSOEP)	Poland	EUSILC Statistics on Income and Living Conditions
Denmark	EUSILC Statistics on Income and Living Conditions	Portugal	EUSILC Statistics on Income and Living Conditions
Dominican Republic	ENFT Encuesta Nacional de Fuerza de Trabajo	Paraguay	EPH Encuesta Permanente de Hogares
Ecuador	ENEMDU Encuesta Nacional de Empleo, Desempleo ...	Romania	EUSILC Statistics on Income and Living Conditions
Egypt	LFSS Labour force sample survey	Rwanda	EICV Integrated Household Survey on Living Conditions
Spain	EUSILC Statistics on Income and Living Conditions	El Salvador	EHPM Encuesta de Hogares de Propósitos Múltiples
Estonia	EUSILC Statistics on Income and Living Conditions	Serbia	EUSILC Statistics on Income and Living Conditions
Ethiopia	NLFS National Labour Force Survey	Slovakia	EUSILC Statistics on Income and Living Conditions
Finland	EUSILC Statistics on Income and Living Conditions	Slovenia	EUSILC Statistics on Income and Living Conditions
Fiji	EUS Employment, Unemployment Survey	Sweden	EUSILC Statistics on Income and Living Conditions
France	EUSILC Statistics on Income and Living Conditions	Thailand	LFS Labour force survey
United Kingdom	EUSILC Statistics on Income and Living Conditions	Tajikistan	LSS Living Standards Survey
Ghana	LFS Labour Force Survey	Timor-Leste	LFS Labour Force Survey
Ghana	LSS Living Standards Survey	Trinidad and Tobago	CSSP Continuous Sample Survey of the Population
Gambia	LFS Labour force survey	Turkey	LFS Household Labour Force Survey
Greece	EUSILC Statistics on Income and Living Conditions	Tanzania, United...	LFS Labour Force Survey
Guatemala	ENEI Encuesta Nacional de Empleo e Ingresos	Uganda	LFS National Labour Force
Guyana	LFS Labour Force Survey	Uruguay	ECH Encuesta Continua de Hogares
Honduras	EHPM Encuesta Permanente de Hogares de Propósitos ...	United States	CPS Current Population Survey
Croatia	EUSILC Statistics on Income and Living Conditions	Venezuela, Bolivarian	EHM Encuesta de Hogares por Muestreo
Hungary	EUSILC Statistics on Income and Living Conditions	Viet Nam	LFS Labour force survey
Indonesia	LFS National Labour Force Survey	Samoa	LFS Labour force survey
India	NSS Employment-Unemployment Survey	Yemen	LFS Labour Force Survey
Iceland	EUSILC Statistics on Income and Living Conditions	Zambia	LFS Labour Force Survey
Italy	EUSILC Statistics on Income and Living Conditions		