

Global inequality: how large is the effect of top incomes?

Vanesa Jordá^(a)

Miguel Niño-Zarazúa^(b)

^(a) Corresponding author: jordav@unican.es. Department of Economics, University of Cantabria. Avda. de los Castros, s/n, 39005, Santander (Spain). Tel. +34 924202275, Fax. +34 942201603.

^(b) miguel@wider.unu.edu, United Nations University - World Institute for Development Economics Research (UNU-WIDER), Katajanokanlaituri 6 B, FI-00160 Helsinki, Finland.

Abstract

Despite the growing interest in global inequality, assessing inequality trends is a major challenge because individual data on income or consumption is not often available. Nevertheless, the periodic release of certain summary statistics of the income distribution has become increasingly common. Hence, grouped data in form of income shares have been conventionally used to construct inequality trends based on lower bound approximations of inequality measures. This approach introduces two potential sources of measurement error: first, these estimates are constructed under the assumption of equality of incomes within income shares; second, the highest income earners are not included in the household surveys from which grouped data is obtained. In this paper, we propose to deploy a flexible parametric model, which addresses these two issues in order to obtain a reliable representation of the income distribution and accurate estimates of inequality measures. This methodology is used to estimate the recent evolution of global interpersonal inequality from 1990 to 2015 and to examine the effect of survey under-coverage of top incomes on the level and direction of global inequality. Overall, we find that item non-response at the top of the distribution substantially biases global inequality estimates, but, more importantly, it might also affect the direction of the trends.

1 Introduction

Over the past two decades, there has been a growing interest in the economic literature and international policy fora in the levels of, and the trends in, global inequality. The UN System Task Team report that preceded the introduction of the Sustainable Development Goal 10, pointed out that “[global] inequality is a key concern, not just from the perspective of a future in which a decent and secure wellbeing is a prerogative of all citizens, but sustained development itself is impeded by high inequalities.”¹ Hence, redressing these trends will be a major challenge in the decades ahead.

When individual records on personal or household income data are available, the estimation of income inequality is relatively simple. However, much of the existing scholarship on economic inequality has been plagued by a lack of individual data. This potential limitation is particularly severe for studies with large geographic coverage, which involve several countries at different points in time. Nevertheless, the periodic release of certain summary statistics of income distribution has become relatively common. The World Income Inequality Database (WIID) and World Bank’s PovcalNet are examples of extensive databases that store grouped income/consumption data. Due to their large geographical coverage, most empirical research on global inequality has relied on grouped data to estimate global and regional trends in income inequality (Bourguignon and Morrison, 2002; Lakner and Milanovic, 2016; Niño-Zarazua et al., 2017; Dowrick and Akmal, 2005; Anand and Segal, 2017). While these studies, for the most part, do point towards virtually identical inequality trends, the methodological weaknesses of most of the existing analyses require a careful interpretation of their results.

A common limitation in much of this research relates to the assumptions made

about the shape of the distribution of income. Grouped data are usually available in the form of few income and population shares, i.e. points of the Lorenz curve. To estimate inequality measures, it is therefore essential to define a method to link such points. With very few exceptions, the extant scholarship on the world income distribution has relied on linear interpolation to approximate the shape of the Lorenz curve from which relative inequality measures can be estimated. The broad popularity of this methodology is not only due to its simplicity, but also because it is argued that there is no need to impose any particular model to fit the empirical data. However, although not explicitly, this approach rests on a predefined distributional model, which assumes that all individuals within a particular quantile have the same level of income. Hence, relative inequality measures estimated within this framework are generally regarded as lower bounds (see, e.g., Kakwani, 1980), which yield severely biased estimates of inequality levels (Jorda et al., 2018).

A second potential source of bias is caused by under-coverage of top incomes in the household surveys from which income shares are generated. To better capture the upper tail of the income distribution, a growing body of studies has used administrative records on personal income tax returns (Atkinson et al., 2011; Piketty and Saez, 2013; Leigh, 2007; Alvaredo et al., 2013).² This renewed interest on the top incomes literature has lead to important developments in data generation, notably the World Wealth and Income Database (WID), which includes series of top income shares from tax records (Alvaredo et al., 2015). Tax data, however, do not provide a reliable representation of the bottom of the income distribution. Hence, tax-based inequality estimates would also be biased downward due to the under-coverage of the poorest individuals.

In this study, we aim to overcome the limitations in the existing literature on the

estimation of the world income distribution to provide accurate estimates of regional and global inequality trends. We develop a new methodology based on parametric models that helps us to mitigate the two sources of bias advanced above. Instead of assuming equality of incomes within shares, parametric distributions define more plausible assumptions about the shape of the Lorenz curve. The parametric approach involves the choice of a functional form that leads to a reliable representation of the income distribution.³ This is a daunting challenge because the analysis of global inequality involves a highly heterogeneous sample of countries in terms of income dynamics. To avoid misspecification bias, we adopt a well-suited functional form, the so-called generalised beta distribution of the second kind (GB2), which nests the parametric assumptions in the literature (see McDonald, 1984, Jenkins, 2009). Although this model provides an excellent fit to income data across different periods and countries (Jorda et al., 2018), this is, to our knowledge, the first study that adopts such a general model to fit the global income distribution.⁴

The methodology developed in this study also takes into account the lower rate of response of the rich in the estimation of income inequality. To deal with this data limitation, we consider the truncated nature of the survey data. Put simply, household surveys may only be representative for the poorest t per cent of the population. Our estimation strategy relies on the truncated distribution to derive the income distribution of the whole population, including a reliable representation of the upper tail. As a result, our approach leads to inequality estimates that incorporate differences across the full income range. This methodological strategy has been already suggested by Anand and Segal (2008): “a possible route [to overcome the undersampling and underreporting problems of top incomes in household surveys] may be to estimate parametrically within-country distributions [...] One could specify a distribution for each country that incorporates a

plausible upper tail and estimate it from household survey data. The estimated distribution would then provide us with corrected estimates for both average income and the level of inequality.”

The main challenge to implement our methodology is the definition of the truncation point t . Recent developments in the definition of optimal truncation levels require the use of individual data to be implemented (Diaz-Bazan, 2015). Although this potential limitation has also been faced by the extant studies in global inequality, the top income group has rarely been optimally chosen. Indeed, most studies set an arbitrary threshold ranged from 1 per cent (Anand and Segal, 2015; 2017) to 10 per cent (Lakner and Milanovic, 2016). While country case evaluations suggest a truncation level of 1 per cent (Jenkins, 2017; Burkhauser et al., 2017), this evidence is mostly based on developed economies. Our analysis, however, also includes developing countries, for which the level of truncation remains unknown. Hence, we make a conservative assumption of setting the maximum level of truncation equal to 1 per cent and perform a sensitivity analysis to check the robustness of the results. In line with prior research, our results point towards a downward trend of global inequality from 1990 to 2015. However, we find that the under-coverage of the richest individuals in household surveys generates a severe bias in global inequality estimates, up to 40 per cent for certain inequality measures. More remarkable, still, is the possible reversal of this decreasing trend if, as expected, non-response rates increase over time.

In the following section, we explore major measurement issues in the estimation of global inequality. The subsequent section discusses the fully-parametric methodology proposed in this study to approximate the Lorenz curve for the entire income distribution. Thereafter, we present the data used to estimate the global income distribution

and provide a thorough description of the selection procedure. Before reporting and discussing the results, we evaluate the goodness-of-fit of our model. The paper concludes with an assessment of the political and practical implications of our findings.

2 Measuring global income inequality

Notwithstanding the significant expansion in the generation of household surveys over the past 40 years, only data on per capita income (or consumption expenditure) are currently available for a significant number of countries over a reasonably long period of time.⁵ The availability of large datasets, such as the WIID and PovcalNet, has motivated the use of grouped data for the analysis of distributional patterns. Most prior research on the world distribution of income has used income shares (typically five or ten points of the Lorenz curve) to estimate global interpersonal inequality. However, the bulk of empirical work has relied mainly on nonparametric methodologies which assume that all individuals within each income share have the same income. These estimates are regarded as being downward-biased estimates of the actual level of global inequality because inequality within income shares is not considered (Bourguignon and Morrison 2002; Milanovic, 2011; Lakner and Milanovic, 2016; Niño-Zarazua et al., 2017; Dowrick and Akmal, 2005).

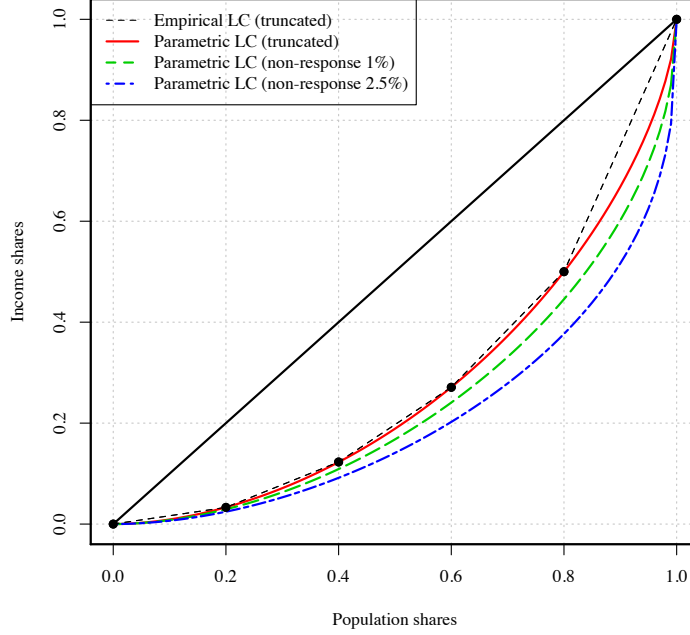
To illustrate this, consider the black points in Figure 1, which are the income shares for the United States in 2013 retrieved from the WIID. These points of the Lorenz curve are, in most cases, the sole information available on the within-country distribution. Thus, to construct the Lorenz curve, we should define a method to link these points. An intuitive approximation would be to interpolate the observed income shares linearly,

as illustrated by the dashed line in Figure 1. One major drawback of using linear interpolation is that these comparisons would be somewhat crude in that we assume that all individuals classified in a given population group have the same income. Although simple, this approach comes at the price of neglecting differences within income shares, leading to downward-biased estimates of inequality. In this case, the dashed Lorenz curve yields a Gini index of 0.442, while the actual Gini index reported by survey data is 0.464. Thus, we need to define a model which allows us to impose more plausible assumptions on income dynamics within income shares to obtain reliable estimates of inequality.

Parametric models are a sound statistical method to estimate inequality from grouped data. The use of a parametric model aims to define a more reliable approximation of the shape of the Lorenz curve between the observed income shares than a rough linear interpolation. Although parametric models seem to be a suitable alternative to nonparametric techniques for estimating income distributions (Dhongde and Minoiu, 2013), very few previous studies have relied on this methodology to estimate income inequality. The reason seems to be the need to make ex-ante assumptions on the shape of the distribution. If our choice is not a representative model of the distribution of income, our estimates of inequality measures would be affected by misspecification bias. Notwithstanding this potential limitation, the parametric approach provides much more accurate results than the conventionally used lower bound (Jorda et al., 2018). Figure 1 confirms this result. The solid red line is the estimated Lorenz curve using the flexible parametric model that we deploy in this study: the GB2 distribution. The estimated Gini index under our parametric assumption is 0.461, which is considerably closer to the actual survey value.

A major drawback of using survey data for estimating national and global inequality

Figure 1: Truncated and non-truncated Lorenz curves under different methodological assumptions



is that the richest households respond to survey questionnaires proportionately less than the rest of the population, resulting in under-coverage of top incomes. This explains why Alvaredo (2009a) could not find rich individuals reporting incomes over one million dollars in Argentina despite the fact that there were about 700 people with such income levels according to tax records.⁶ Burkhauser et al. (2017) identify two main sources of under-coverage of top incomes in household surveys. The first source of under-coverage arises from under-reporting of income among the richest individuals. Secondly, there may be no respondents at all from the extreme right tail, because the survey design does not target high-income earners or because the richest individuals refuse to participate. Both types of under-coverage contribute to downward bias survey estimates of inequality because there is not enough income observed at the very top income range.

A small but growing number of studies have aimed to estimate inequality measures

that incorporate differences across the full income range. The analytical methods that account for the bias arising from under-coverage of top incomes can be categorised into two main types (see Jenkins, 2017): those that extrapolate the upper tail using only survey income data; and those that use tax records to approximate the top tail of the income distribution. The first approach estimates inequality of the poorest t per cent of the distribution using survey data and derives an inequality estimate of the $1 - t$ per cent of the richest by fitting a Pareto-type distribution (Atkinson, 2007; Alvaredo and Londoño-Vélez, 2011). This method addresses the issue of unit non-response among the rich, but the problem of under-reporting remains unsolved. Although estimates of top income shares from this semi-parametric approach seem to be lower than those obtained from tax records, this methodology “improves the performance of those inequality indices that are normally considered particularly sensitive to extreme values” (Cowell and Flachaire, 2007). The second approach is virtually identical to the former one, but uses tax records to estimate the parametric model at the top of the income distribution. Because this method draws on tax data as an external reference point, it addresses both under-reporting and unit non-response.

Although country-case studies are becoming increasingly common, scant research has explored the impact of under-coverage of the richest individuals at the global level. Lakner and Milanovic (2016) corrected the lower-bound estimates of global inequality by fitting a Pareto distribution to the top 10 per cent of the distribution. Survey data were adjusted by the gap between national accounts and survey income, as a proxy for the extent of under-coverage. While this semi-parametric technique represents an important step towards an accurate estimation of global inequality, it presents three potential limitations. First, these estimates ignore differences within income shares.

Second, it is rather arbitrary to allocate all the excess of national accounts to survey means at the top decile. Empirical analyses in industrialised countries suggest that under-coverage is not an issue for the bottom 95 per cent of the distribution (Jenkins, 2017). Hence, there is no empirical or theoretical justification to chose that particular threshold. Finally, the validity of the Pareto distribution to represent income dynamics at top of the distribution has been questioned (Atkinson, 2017; Blanchet et al., 2017; Jenkins, 2017).

To better capture income dynamics at the top of the distribution, Anand and Segal (2015; 2017) and Hong et al. (2019) combined tax data from the WID and household surveys from Povcalnet and the WIID to evaluate the evolution of global inequality. A major limitation of this approach to estimate global inequality is that tax data is too scarce, especially among developing countries, to grant a comprehensive adjustment for survey under-coverage of top incomes. Another source of measurement error arises from comparability issues across data sources. Survey data and tax records should refer to the same income definition and population. Surveys typically contain information on household incomes, whereas the individual is the unit of analysis in tax records. Moreover, tax data refer to individuals aged 15 years or more, but survey data cover the whole population. Finally, some countries and repositories, such as Povcalnet, mainly report consumption data, which poses an additional comparability issue.

Prior studies that have attempted to correct survey-based estimates using tax data have emphasised the need for reconciling survey and tax data definitions. Analyses that merge these two data sources are prone to the “apples and oranges” comparability problem, which would certainly introduce an additional source of measurement error (Jenkins, 2017, Burkhauser et al., 2017). Because global evaluations of income inequality

rely on grouped data, it is not possible to reconcile the income definitions of household surveys and tax records. We argue that, in the absence of comprehensive publicly available data on individual incomes, under-coverage issues must be addressed using only survey data. Hence, although survey under-coverage of top incomes might be the result of under-reporting among the very rich and unit non-response, only this last issue can be addressed with grouped data.

Household survey data are representative of the bottom t per cent of the population, whose Lorenz curve is represented by the solid red line in Figure 1. In more formal terms, this is equivalent to estimate the Lorenz curve of the distribution of income conditional on being below a particular threshold, i.e. $f(y|y < b)$, where b is the minimum income of the top $(1 - t)$ per cent of the richest individuals. Thus, survey income data are right-truncated samples of the income distribution. Our interest, however, resides in producing estimates of global inequality for the entire population. The income distribution $f(y)$ can be easily obtained from the conditional distribution as follows:

$$f(y) = f(y|y < b)F(b),$$

where $F(b) = t$ is the cumulative distribution function (cdf) evaluated at the truncation point, which reports the proportion of the population covered by the survey.

The green and the blue lines in Figure 1 represent the Lorenz curves of the unconditional distribution ($f(y)$) assuming that the survey covers 99 per cent ($1 - t = 0.01$) and 97.5 per cent ($1 - t = 0.025$) of the population respectively. The truncated distribution (red) Lorenz dominates the unconditional distributions because top incomes are not included in the sample, thus being characterised by lower levels of inequality. It is

quite intuitive that, as the proportion of the population covered by the survey increases, the effect of truncation on the entire population diminishes. In the limit, b equals to the income of the richest individual, so that $F(b) = 1$ and then $f(y) = f(y|y < b)$. The effect of missing top incomes on global inequality can be considerable even when household surveys cover a large proportion of the population. In our example, if the survey covered 99 per cent of the population, the Gini index of the unconditional distribution would be 0.521, whereas, for a coverage rate of 97.5 per cent, it would rise to 0.592.

3 Methodology

In this study, we propose a fully-parametric approach to approximate the Lorenz curve of the entire income distribution for each country and year. Among the whole range of models proposed for the size distribution of income, the GB2 family seems to be the most appealing option. It is a general class of distributions that is acknowledged to provide an accurate fit to income data (McDonald and Xu, 1995; McDonald and Mantrala, 1995; Jorda et al., 2018). Moreover, this family nests most of the functional forms used to model the size distribution of income (Kleiber and Koltz, 2003), including the Beta 2 distribution (Chotikapanich et al., 2012), the lognormal and the Weibull distributions (Pinkovskiy and Sala-i-Martin, 2014; Chotikapanich et al., 1998) and the Lamé family (Jorda et al., 2014). Hence, it would converge to any of these models if needed.⁷ Even though the same functional form is fitted to all countries over the whole period, the parameters of the GB2 vary across countries and over time, thus allowing us to model the idiosyncrasy of each country.

The GB2 distribution is defined in terms of the probability density function (pdf)

$(a, \beta, p, q \geq 0)$ as (McDonald, 1984):

$$f(x; a, \beta, p, q) = \frac{ax^{ap-1}}{\beta^{ap}B(p, q)[1 + (x/\beta)^a]^{p+q}}, \quad x \geq 0,$$

where $B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1}dt$ is the beta function. The parameters a, p and q are shape parameters and β is a scale parameter.

Since our estimation strategy relies on points of the Lorenz curve, we need to define it for the GB2 distribution. Following Chotikapanich et al. (2018) and Arnold and Sarabia (2018), the Lorenz curve of the GB2 family is given by,

$$L_{GB2}(u; a, p, q) = B\left(B^{-1}(u; p, q); p + \frac{1}{a}, q - \frac{1}{a}\right), \quad 0 \leq u \leq 1, \quad (1)$$

where $q > 1/a$ and $B^{-1}(x; p, q)$ is the inverse of the incomplete beta function ratio given by $B(v; p, q) = \int_0^v t^{p-1}(1-t)^{q-1} dt / B(p, q)$.

If we estimated directly Eq. (1) using survey data, which only include information about the bottom t per cent of the population, we would estimate the parameters of the truncated distribution ($f(y|y < b)$). Hence, survey under-coverage of top incomes would not be addressed. To consider the right truncation of survey data, we estimate the following model:

$$L(u|u < t) = \frac{L(u)}{L(t)}, \quad (2)$$

where $L(u)$ is the Lorenz curve of the entire population, $t \in [0, 1]$ is the proportion of the total population covered by the survey (defined as $F(b)$ in terms of the cdf), so $L(t)$ is the Lorenz curve at the truncation point, i.e. the share of the total income held by

the population covered in the survey. Substituting the formula of the Lorenz curve of the GB2 distribution (Eq. 1) in Eq. (2) we obtain,

$$L_t(u; a, p, q) = \frac{B(B^{-1}(u; p, q); p + \frac{1}{a}, q - \frac{1}{a})}{B(B^{-1}(t; p, q); p + \frac{1}{a}, q - \frac{1}{a})}. \quad (3)$$

The parameters of the distribution are estimated by minimizing the squared deviations between the income shares and the theoretical points of the truncated Lorenz curve of the GB2 distribution given in Eq. (3), that is

$$\min_{a, p, q} \sum_{j=1}^J \left(\frac{B(B^{-1}(u; p, q); p + \frac{1}{a}, q - \frac{1}{a})}{B(B^{-1}(t; p, q); p + \frac{1}{a}, q - \frac{1}{a})} - s_j \right)^2. \quad (4)$$

Although the parameters are estimated from a truncated Lorenz curve, the estimates belong to the distribution of the whole population, which also includes the omitted top incomes. Hence, we can obtain the Lorenz curve of the whole population just by substituting them in Eq. (1).

The β parameter plays no role in the estimation because the Lorenz curve is independent to scale. To estimate the scale parameter, we equal the theoretical expression of the mean of the GB2 distribution to an estimate of per capita income and solve it for the β parameter:

$$\hat{\beta} = \bar{Y} \frac{B(\hat{p}, \hat{q})}{B(\hat{p} + \frac{1}{\hat{a}}, \hat{q} - \frac{1}{\hat{a}})}, \quad (5)$$

where \bar{Y} denotes the per capita income, $\hat{a}, \hat{p}, \hat{q}$ are the parameters estimated using Eq. (4) and $B(.,.)$ stands for the beta function.

It is important to highlight here that, due to the truncated nature of survey data,

t is not a parameter to be estimated. Hence, the proportion of the population covered by household surveys must be defined before the estimation of Eq. (4). Although prior research has made different assumptions on the value of this parameter (Anand and Segal, 2015; 2017; Lakner and Milanovic, 2016), there is not a universal truncation point and non-response rates are expected to vary over time. Setting an arbitrary threshold might severely bias the results, thus posing a major challenge to obtain reliable estimates. Country case evaluations in developed countries suggest that survey non-response is not an issue for the bottom 99 per cent of the distribution (Jenkins, 2017; Burkhauser et al., 2017). Our analysis, however, also involves developing countries for which the non-response rate $(1 - t)$ is expected to be lower than 1 per cent. The reason seems to be that rich individuals represent a much lower proportion of the population in developing countries (Anand and Segal, 2017).

To a greater or lesser extent, non-response seems to be systematically observed in virtually all household surveys. The definition of the truncation point is particularly problematic if our choice of $1 - t$ is in some countries above the actual rate of non-response while in others is below that rate. If the actual proportion of the population not covered by the survey in a particular country is smaller than our truncation parameter, our estimates of income inequality will be downward-biased because, by definition, distributions with lower truncations points (t) are Lorenz dominated by the distributions with higher truncation points. On the contrary, if we set t below the actual level of the population covered by the survey, meaning that non-response rates are overstated, inequality will be biased upwards. Hence, we would not be able to determine the direction of the bias. We argue, therefore, that the choice of $1 - t$ should be sufficiently small to ensure that it is below the actual non-response rate of all surveys so that our estimates would be char-

acterised as lower bounds of global inequality. Thus, we make a conservative assumption of setting the maximum level of non-response to 1 per cent and provide a battery of estimates using different assumptions below this level.

4 Data

For the analysis of global income inequality, we use data on income shares from the WIID version 3.4 (UNU-WIDER, 2017), which contains information on Gini coefficients and income (or consumption) shares for 182 countries over the period 1867-2015.⁸ The WIID is the most reliable and comprehensive database of worldwide distributional data currently available.⁹ Our analysis focuses on the period 1990-2015 at five-year intervals – 1990, 1995, 2000, 2005, 2010 and 2015. Whenever we had missing data for the exact year, we opted to include observations within a maximum of the previous/next five years of each data point, although preference was given naturally to the closest observations.

In addition, we adopt a conceptual base of the Canberra Group to minimise the problems that may arise from conceptual differences in the WIID in terms of the unit of analysis, the equivalence scale, the quality of the data and the welfare concept. First, as we focus on global interpersonal inequality, the preferred unit of analysis is the individual rather than the household. Second, we opt for income per capita rather than adult equivalent adjustments. Third, we give preference to observations from nationally representative surveys, which are deemed to be of the highest quality in the WIID. Finally, in relation to the welfare concept, our preference is to choose income-based data instead of consumption-based data. However, dropping consumption-based data altogether would have severely affected the coverage of the global population. In order to keep the global

coverage as high as possible, we also include consumption data, covering around 90 per cent of the global population in all years (see Appendix Table A1).

Mixing consumption and income data could potentially bias the results because both variables present different distributional patterns, being consumption typically characterised by lower inequality levels. Hence, consumption observations need to be adjusted to harmonise, at least partially, both kinds of data. In this study, the homogenisation procedure consists of comparing the average income shares with those of consumption, for the available datasets that have both income and consumption data available for the same country and year. If there are different sources for income and consumption data for a given country-year, our preference is to choose instances where both kinds of data came from the same sources. This is done in order to minimise measurement error due to variations in survey designs. We group countries into eight world regions and compute the average index of income relative to consumption (see Appendix Tables A2 and A3). Previous studies have used the absolute average difference between both welfare concepts to correct either consumption shares (Niño-Zarazúa et al., 2017) or Gini indices (Deininger and Squire, 1996). We opt, however, for the relative difference at the regional, not the global, level to better account for the heterogeneity of countries in the income-consumption relationship.

To construct the global distribution of income, we need in addition to income shares, data on mean income. The choice between mean incomes from national accounts or household surveys is generally a key question in the analysis of global inequality. With a few exceptions (notably Anand and Segal, 2015; Milanovic, 2011, Lakner and Milanovic, 2016), most studies on global inequality have used national accounts, and in particular gross domestic product (GDP) per capita, due to the limited availability of survey

means (Atkinson and Brandolini, 2010; Bhalla, 2002; Bourguignon and Morrisson, 2002; Dowrick and Akmal, 2005; Niño-Zarazúa et al., 2017; Sala-i-Martin, 2006; Jorda et al., 2014; Chotikapanich et al., 2012). It is worth noting though that, for the specific objective of this study, which aims to account for the effect of omitted top incomes on global inequality, the use of mean incomes from household surveys would yield biased estimates of global inequality given the persistent under-coverage of incomes at the upper tail of the distribution (Anand and Segal, 2008).

Since data from national accounts may actually be a better proxy for the actual average income level, we use GDP per capita adjusted by purchasing power parities (PPP) at constant prices of 2011, taken from the World Bank’s World Development Indicators. We are aware that discrepancies between national accounts and survey means are not only driven by the undersampling and underreporting of top incomes. Deaton (2005) argues that GDP per capita is a poor measure of household income as it contains depreciation, retained earnings of corporations, and components of government revenue that are not distributed back to households in the form of social assistance or social security transfers. Though not a perfect indicator, GDP per capita is nevertheless the best alternative among the available measures of mean income, which additionally allows for further comparison with most previous studies.

5 Goodness-of-fit

How robust are our results? A potential limitation of using parametric models to estimate income inequality is the requirement to impose a particular functional form to describe the income distribution that could lead to biased estimates in case that the model is

not able to represent adequately the shape of the income distribution and/or the Lorenz curve. Since the validity of our estimates relies on the assumption that the income variable follows a GB2 distribution, assessing the goodness-of-fit (GOF) of our model is fundamental. Jorda et al. (2018) provide a thorough examination of the implications of using the GB2 family to estimate income inequality from 5570 datasets of grouped data. The GB2 distribution is confirmed as the best candidate to estimate income distributions, which leads to more reliable estimates of inequality measures than the lower bound approximation in virtually all cases.¹⁰

Despite the excellent performance of the GB2 distribution to estimate income inequality from grouped data, income shares are obtained from truncated samples. Since one of the main contributions of this study is to address non-response issues in household surveys, GOF evaluations should also focus on the performance of the proposed method to obtain accurate estimates of inequality measures and a reliable representation of the incomes at the upper tail of the distribution. To do so, we rely on data from the Luxembourg Income Study (LIS). The LIS database gathers harmonised micro-data on disposable income collected for nearly 50 countries, over the period from 1980 to 2016. Using the 292 datasets of individual records available in the ten waves of the LIS database, we compute the Gini index and the income share accruing to the 5 per cent and the 1 per cent of the richest individuals in the surveys. These statistics have been obtained following the methodological guidelines of LIS.¹¹ In our GOF exercise, these estimates are taken as the inequality indicators of the whole population. We then remove the top 1 per cent of the richest individuals of household samples to artificially generate item non-response at the top of the distribution. For these truncated samples, we reconstruct grouped data with the same structure as the WIID: ten income shares

Table 1: Absolute error in the estimation of the Gini index and top incomes shares from truncated data

Truncation point	Mean	10%	20%	40%	60%	80%	90%
Gini index							
$t = 0.99$	0.0072	0.0007	0.0017	0.00380	0.0070	0.0104	0.0155
$t = 0.9925$	0.0098	0.0035	0.0044	0.00670	0.0096	0.0134	0.0178
$t = 0.995$	0.0138	0.0066	0.0082	0.01050	0.0135	0.0176	0.0221
$t = 0.9975$	0.0185	0.0099	0.0119	0.01500	0.0185	0.0228	0.0297
$t = 1$	0.0244	0.0134	0.0157	0.01970	0.0234	0.0313	0.0395
Share of total income held by the richest 5%							
$t = 0.99$	0.0100	0.0009	0.0020	0.0053	0.0087	0.0147	0.0215
$t = 0.9925$	0.0115	0.0021	0.0038	0.0070	0.0109	0.0170	0.0236
$t = 0.995$	0.0149	0.0045	0.0069	0.0106	0.0144	0.0209	0.0266
$t = 0.9975$	0.0194	0.0075	0.0102	0.0142	0.0186	0.0250	0.0345
$t = 1$	0.0255	0.0108	0.0138	0.0185	0.0238	0.0330	0.0456
Share of total income held by the richest 1%							
$t = 0.99$	0.0104	0.0007	0.0018	0.0048	0.0086	0.0144	0.0220
$t = 0.9925$	0.0107	0.0014	0.0026	0.0056	0.0094	0.0150	0.0231
$t = 0.995$	0.0120	0.0022	0.0040	0.0074	0.0111	0.0163	0.0238
$t = 0.9975$	0.0141	0.0039	0.0055	0.0089	0.0134	0.0199	0.0262
$t = 1$	0.0174	0.0058	0.0077	0.0117	0.0164	0.0227	0.0334

Results based on 292 datasets of the LIS database. Parametric models have been estimated by NLS using Eq. (4). The Gini index has been estimated by Monte Carlo simulation of synthetic samples of size $N = 10^6$.

and the Gini coefficient, which is used to deploy the estimation procedure developed in this paper.

As a first measure of GOF, we have computed the difference between the Gini index obtained from the whole sample of the survey with the Gini index estimated from grouped data of the simulated truncated sample. To replicate the estimation of income inequality performed in the next section, we set $t = 0.99, 0.9925, 0.995, 0.9975$ and 1 in the estimation of Eq. (4). A summary of this information is shown in Table 1, which presents the deciles of the difference between the Gini index of the whole sample and the Gini index estimated from truncated grouped data. Since the simulated rate of non-response is 1 per cent, the truncation point $t = 0.99$ leads to the most accurate results.

Our estimates indicate that, in 80 per cent of the cases, the gap between the Gini index of the whole population and the estimated one is lower than 0.01 points. As the truncation point increases, the size of the bias in the estimation of the Gini index also rises. Setting t equal to one means that truncation is not considered in the estimation. Inequality estimates that do not consider the truncation of the data present Gini indices between 1 and 5 points below the “true” Gini index.

These preliminary results suggest that the novel approach proposed in this paper provides reliable estimates of the Gini index if the truncation point (t) is set close to the actual proportion of the population covered by the survey. Turning now our attention to the estimates of the top income shares, we evaluate the performance of the estimation method to represent the upper tail of the income distribution. On average, for a truncation point of 99 per cent, the absolute error in the estimation of the share of the richest 5 per cent is about 1 per cent. Although for this level of truncation, the estimation errors of the share of the richest 1 and 5 per cent are of similar size, when non-response rates are understated, the error seems to be larger for the share of the top 5 per cent.

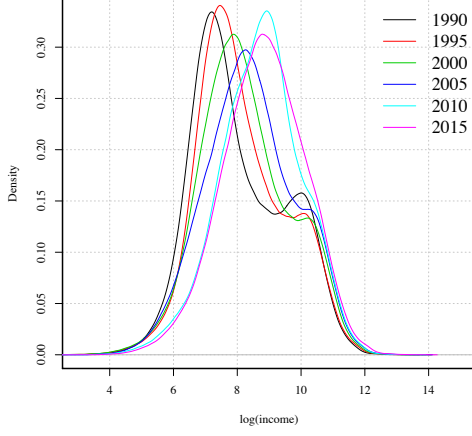
6 Results

Before moving onto the estimates of global inequality, we focus first on the evolution of the global distribution of income under different assumptions on the non-response rate. The first panel of Figure 2 depicts these estimates without considering survey under-coverage of top incomes, using a mixture of GB2 distributions. During the 1990s, the global distribution of income presents two different modes, the main one around 9000\$ and a smaller peak of richer individuals around 20000\$. By 2015, the global distribution

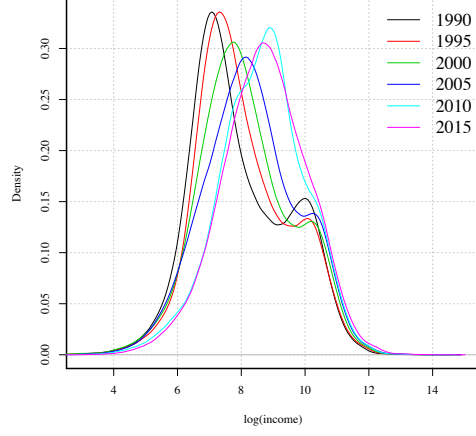
of income becomes bell-shaped, with the mode observed at 6000\$. This change in the distribution from bimodal to unimodal might be partially explained by the evolution of average income levels. Indeed, the distribution of GDP per capita, which traditionally presented two well-defined modes of poor and rich countries, has experienced a similar evolution. Both peaks are becoming gradually less pronounced since 2000 due to the progress in average income levels of some countries from the poor mode that have moved towards the rich one (Krause, 2017). This outstanding progress in GDP per capita of large countries, especially China, has also led to a rightward movement of the global income distribution (Lakner and Milanovic, 2016).

Looking now at the other panels of Figure 2, our estimates suggest that, although under-coverage is an issue at the top of the income distribution, the effect of considering the richest individuals in the estimation is also evident at the bottom tail. As the truncation point (t) falls, meaning that the non-response rate increases, a greater proportion of individuals is observed at the left tail. This distributional shift is caused by the method used to estimate the scale parameter of national distributions, which makes the mean of the distribution equal to the GDP per capita (see Eq. (5)), independently of the assumed level of truncation. Therefore, we are estimating inequality in truncated distributions but scaling them to have the same mean as the whole distribution. Because National Accounts are not affected by under-coverage issues at the top of the distribution, this method leads to reliable estimates of per capita income. However, to impute the top incomes not considered in the survey while keeping the mean constant, incomes at all tranches of the distribution must be reduced. This assumption has no impact on national inequality levels because relative inequality measures are scale independent. At the global level, this approach leads to more accurate estimates than

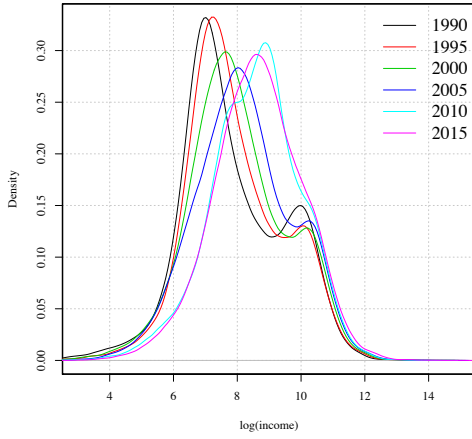
Figure 2: Evolution of the global distribution of income over time, logarithmic scale.



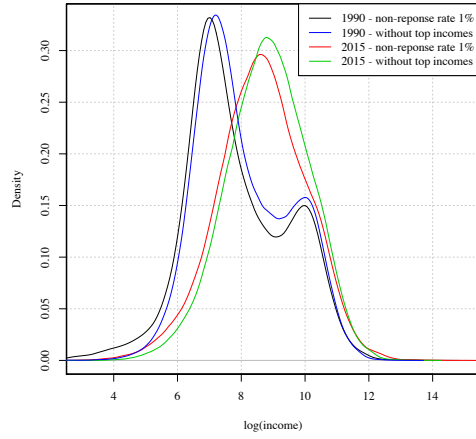
(a) Without top incomes ($t = 1$)



(b) Truncation point $t = 0.995$



(c) Truncation point $t = 0.99$



(d) $t = 1$ vs $t = 0.99$

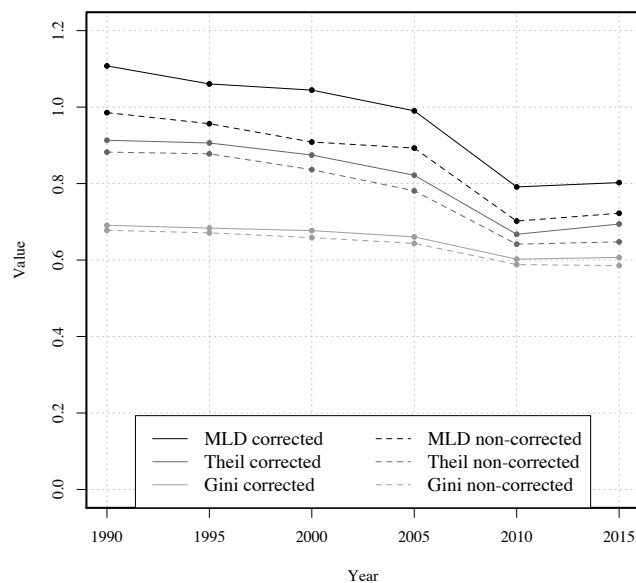
scaling distributions using the truncated (survey) mean income. Since our estimates of per capita income are robust to under-coverage issues, we can produce accurate estimates of between-country inequality. Hence, biases on global inequality can only be introduced by the within-country component.

We focus now on the evolution of global income inequality from 1990 to 2015. We present estimates of the Gini index, which seems to be the most popular measure due to its intuitive interpretation in terms of the area between the Lorenz curve and the

egalitarian line. This inequality measure is sensitive to the middle of the distribution, but it does not allow us to change the weight given to differences at specific parts of the distribution.¹² In order to vary the importance of redistribution movements at different parts of the distribution, we compute a set of alternative inequality indices belonging to the generalised entropy (GE) family. This family of inequality measures is additively decomposable in two components, the between- and within-country components, and includes a sensitivity parameter that gives weights to differences observed across the income distribution. The mean log deviation (MLD) corresponds to the GE measure when the parameter is set to 0, thus being more sensitive to the bottom part of the distribution. The case given by the Theil's entropy measure, characterised by a parameter value equal to 1, is equally sensitive to all parts of the distribution.

We begin by evaluating the effect of correction of consumption shares in our estimates of global inequality. Figure 3 presents the evolution of the MLD, the Theil index and the Gini coefficient before and after correcting for consumption shares. Corrected and non-corrected series generally show very similar evolutions of income inequality except for the MLD, which presents slightly divergent trends over the period 2000-2005. The parallel evolutions observed in Figure 3 suggest that the correction procedure leads to the expected relationship between income and consumption, with the former being more unequal than the later. If consumption shares are corrected, the Gini index decreased from 0.691 in 1990 to 0.607 in 2015, whereas non-corrected estimates of this inequality measure fell from 0.678 to 0.586. Greater differences between corrected and not corrected estimates are observed for the GE measures, which are considerable for the MLD. One potential explanation would be that, unlike the Gini index, GE measures are not bounded between 0 and 1. Moreover, the MLD is particularly sensitive to the bottom

Figure 3: Evolution of global inequality with corrected and non-corrected consumption shares



tail, which is the part of the distribution most severely affected by the correction factors (see Tables A2 and A3). It should be, however, noted that the income-consumption relationship might be different in each country. Even though our proposal partially captures such heterogeneity by using different correction factors for eight world areas, we are certainly aware that it does not achieve a full reconciliation of consumption and income definitions. Nonetheless, and despite the potential limitations of our method, it is essential to harmonise income and consumption definitions for comparative purposes.

We turn now our attention to the evolution of global inequality. Table 2 presents the results of global inequality estimates using the MLD, the Theil index and the Gini index. To begin the discussion, we focus first on the level of global inequality without taking into account survey under-coverage of top incomes ($t = 1$). These estimates are regarded as a benchmark to assess the bias due to non-response issues at the upper

Table 2: Global income inequality and estimated bias due to survey under-coverage of top incomes

		1990	1995	2000	2005	2010	2015	1990-2000 Change (%)	2000-2015 Change (%)
MLD	Total ($t = 1$)	1.1077	1.0606	1.0445	0.9902	0.7913	0.8026	-5.71	-23.16
	Between	0.7333	0.6328	0.6075	0.5403	0.4425	0.3763	-17.17	-38.06
	Within	0.3744	0.4277	0.4370	0.4499	0.3488	0.4263	16.72	-2.44
Bias (%)	$t = 0.9975$	-11.27	-14.87	-15.23	-12.64	-12.53	-13.40	-1.31	-24.78
	$t = 0.9950$	-19.41	-23.67	-24.41	-21.85	-20.97	-22.44	0.53	-25.12
	$t = 0.9925$	-26.79	-31.34	-32.38	-30.27	-28.56	-30.54	2.07	-25.19
	$t = 0.9900$	-33.36	-38.18	-39.38	-37.89	-35.32	-37.85	3.65	-25.05
Theil Index	Total ($t = 1$)	0.9133	0.9061	0.8746	0.8218	0.6674	0.6941	-4.24	-20.63
	Between	0.6354	0.5913	0.5820	0.5171	0.4083	0.3481	-8.40	-40.19
	Within	0.2780	0.3149	0.2926	0.3047	0.2591	0.3461	5.28	18.26
Bias (%)	$t = 0.9975$	-3.44	-4.46	-4.75	-4.70	-4.39	-5.43	-2.92	-20.07
	$t = 0.9950$	-5.97	-7.53	-7.97	-8.08	-7.46	-9.13	-2.16	-19.62
	$t = 0.9925$	-8.44	-10.41	-11.00	-11.32	-10.36	-12.55	-1.48	-19.22
	$t = 0.9900$	-10.91	-13.24	-13.95	-14.55	-13.17	-15.84	-0.85	-18.85
Gini	Total ($t = 1$)	0.6909	0.6835	0.6770	0.6606	0.6023	0.6069	-2.02	-10.36
Bias (%)	$t = 0.9975$	-1.69	-2.21	-2.24	-2.22	-2.23	-2.79	-1.47	-9.85
	$t = 0.9950$	-2.81	-3.57	-3.67	-3.77	-3.68	-4.68	-1.13	-9.41
	$t = 0.9925$	-3.84	-4.80	-4.99	-5.24	-5.02	-6.42	-0.83	-8.99
	$t = 0.9900$	-4.89	-5.96	-6.26	-6.74	-6.24	-8.08	-0.58	-8.59

Note: The parameters of the GB2 distribution have been estimated by NLS using Eq. (4). Inequality measures have been estimated by Monte Carlo simulation of synthetic samples of size $N = 10^6$.

tail of the distribution. Our estimates reveal a world characterised by extraordinarily high levels of income inequality, even higher than those observed in the most unequal countries. As in earlier studies, all inequality measures do exhibit a declining trend over the past two decades, particularly since 2000.

To further investigate the drivers of such a decrease, we exploit the decomposability property by population subgroups of the GE measures to break down overall inequality into differences in GDP per capita and income disparities within countries. Our estimates suggest that the decrease in global inequality has been largely driven by a decline in between-country inequality, fuelled by the rapid economic growth that populous coun-

tries, such as China and India, have experienced over the past 30 years (Niño-Zarazua et al., 2017; Lakner and Milanovic, 2016). On the contrary, within-country inequality rose during the period 1990 to 2015: from 0.3744 to 0.4263 in the case of the MLD; and from 0.2780 to 0.3461 in the case of the Theil index. As a result, the between-country contribution to global inequality has decreased since 1990. According to both the MLD and the Theil index, differences in mean income between countries accounted for two-thirds of global inequality in 1990; by 2015, the weight of this component on global inequality fell to 0.5.

Thus far, we have conducted a conventional analysis of global inequality without accounting for the effect of omitted top incomes. To estimate of the size of bias in global inequality due to survey under-coverage at the upper tail, we deploy the methodology proposed in Section 3 for different assumptions about the proportion of the population covered by household surveys. Although our methodology allows for different truncations points, we set the same level of truncation in all countries due to the absence of information about the optimal truncation point for every country in the world. Given the rigidity of this assumption, we use non-response rates up to 1 per cent. As discussed earlier, our estimates can be interpreted the minimum level of inequality that would exist if survey data represented the bottom t per cent of the population or less in all countries. Therefore, our results can be regarded as lower bounds of global income inequality assuming that all surveys are affected by non-response rates of, at least, $1 - t = 0.25, 0.5, 0.75, 1$ per cent.

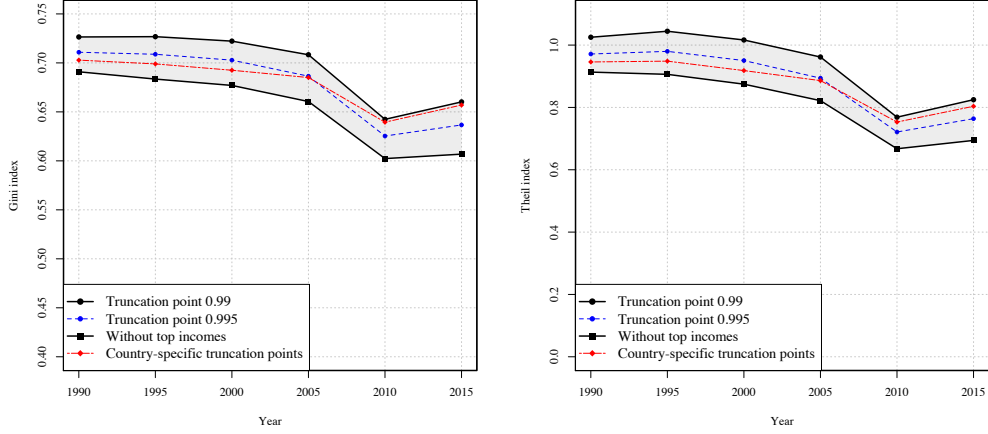
Our estimates suggest that the bias in global inequality estimates is considerable even for the most conservative levels of truncation. The MLD shows the largest bias because it is more sensitive to the bottom of the distribution, which, as observed in

Figure 2, is also affected by the consideration of non-response rates. Naturally, the size of the bias increases as the truncation point decreases. We find an underestimation of global inequality levels in the order of 11 and 39 per cent using the MLD; between 3 and 16 per cent using the Theil index; and between 2 and 8 per cent for the Gini index. Hence, our results show that the effect of survey under-coverage of top incomes can be sizable, but more importantly, that not only the level of global inequality can be affected, also the direction of the trends. This is illustrated in the last two columns of Table 2, which show the growth rate of inequality under different assumptions about the truncation level for the periods 1990-2000 and 2000-2015. For the MLD, the decreasing trend observed during the 1990s becomes positive when a truncation point of 99.5 per cent is considered.

To provide a clearer picture of the patterns of global inequality under different assumptions about the non-response rate, we present the evolution of the Gini index, the MLD and the Theil index in Figure 4. Assuming that non-response rates remain constant over time, the trends of the Theil index and the Gini coefficient do not seem to be affected by the consideration of under-coverage issues in household surveys. These measures exhibit parallel trends that move upwards as the non-response rate $(1 - t)$ increases. As advanced above, the MLD shows different evolutions over the 1990s, depending on the level of truncation. Without modelling the inherent truncation of survey data, this inequality measure shows a decreasing trend from 2000 to 2005. When survey under-coverage is considered in the estimation, the MLD presents an ascending trend over the same period.

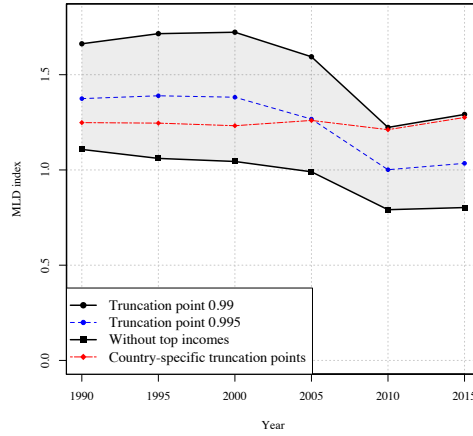
A critical question is how inequality levels and trends are affected by the consideration of different truncation points across countries and over time. Assuming the same

Figure 4: Evolution of global income inequality from 1990 to 2015



(a) Gini index

(b) Theil index



(c) MLD

level of truncation for all countries is highly restrictive and prior empirical evidence from the UK and the US suggests that the issue of under-coverage of top incomes in household surveys has become more severe over the last decades (Jenkins, 2017). Despite the lack of information about country-specific truncation points, our estimates cover a large proportion of the potential scenarios. Assuming that no survey is affected by non-response rates greater than 1 per cent, we can be certain that the actual level of global inequality lies between the estimates that consider this maximum rate of non-response

in all countries and those that do not contemplate the truncation of survey data. This is illustrated by the solid lines in Figure 4. In this sense, our estimates can be interpreted as “inequality bands” that account for the uncertainty about the non-response rate in household surveys. Following this conservative approach, we are only certain that the Gini and the Theil indices fell from 1990 to 2015. The gap between MLD estimates without considering survey under-coverage of top incomes and the estimates assuming a non-response rate of 1 per cent is so wide that no robust conclusions can be drawn about the recent evolution of global inequality.

To further investigate this issue, we construct a counterfactual scenario in which countries in the world regions as defined by the United Nations Development Program (UNDP) present different truncation levels.¹³ More precisely, we assume a non-response rate of 0.25 per cent in all countries over the 1990s; In 2005, the non-response rate rises to 0.5 per cent in Latin America, the two African regions and the two Asian regions; and then from 2010 to 2015, it raises again up to 1 per cent in the same world regions. The evolution of global inequality under this scenario is illustrated by the red lines in Figure 4. The trends of the Theil and the Gini indices do not seem to be strongly affected by the consideration of different truncation points across countries. However, the MLD shows a radically different evolution. The downward trend observed from 2000 to 2010 under the assumption of constant truncation points over time becomes positive in this scenario. As a result, global inequality rises from 1990 to 2015.

Since the consideration of survey under-coverage of top incomes might affect not only the levels of global inequality but also the trends, the findings from prior research should be treated with great caution. The validity of these bands of global inequality relies on the assumption that no survey omits more than the top 1 per cent of the richest

population in any country. Although this might be debatable, prior research suggests that it is conceivable that household surveys do not present non-response rates higher than 1 per cent (Jenkins, 2017; Burkhauser et al., 2017). In the hypothetical case that some surveys included in the analysis presented higher levels of under-coverage, global inequality levels would be higher than those presented in this paper.

The effect of survey under-coverage of top incomes can, nonetheless, vary across world regions, depending on the shape of the truncated distribution, and also the type of welfare institutions, fiscal policies and the social contracts that dominate in different regions of the world. To analyse the heterogeneity in regional distributions, we present inequality estimates based on the MLD for the eight world regions based on the UNDP classification (Table 3).¹⁴ By contrast to what we observe at the global level, regional estimates suggest that within-country inequality plays a major role in regional inequality. This indicates that world regions tend to be, on average, more homogenous in terms of per capita income.

In 1990, East Asia and the Pacific and sub-Saharan Africa were the regions that presented the highest level of inequality. Although inequality in sub-Saharan Africa exhibited a decreasing trend over the period under analysis, it was still the region with the highest income inequality in 2015, reporting an MLD of 0.9197. By contrast, East Asia and the Pacific has seen major reductions in income inequality, largely explained by market-oriented structural reforms, technological change, trade liberalisation and the rapid convergence process that countries such as China, India, Indonesia and Vietnam experienced over the past 30 years vis-à-vis the most advanced economies in the region (see Monfort and Nicolini (2000) and Behrens et al. (2007) for a theoretical discussion).

Table 3: Regional income inequality and estimated bias due to survey under-coverage of top incomes. MLD

		1990	1995	2000	2005	2010	2015
East Asia and the Pacific	MLD ($t = 1$)	0.9419	0.7915	0.7064	0.6225	0.3751	0.4563
	MLD between	0.6447	0.4572	0.3547	0.2482	0.1549	0.0640
	MLD within	0.2972	0.3343	0.3517	0.3743	0.2202	0.3924
	$t = 0.995$	1.0604	0.9754	0.8452	0.7729	0.4219	0.5946
	$t = 0.99$	1.1865	1.1688	0.9957	0.9344	0.4721	0.7532
Europe	MLD ($t = 1$)	0.2714	0.2942	0.2814	0.2709	0.2408	0.2318
	MLD between	0.0834	0.0935	0.0896	0.0652	0.0501	0.0369
	MLD within	0.1880	0.2007	0.1919	0.2056	0.1907	0.1949
	$t = 0.995$	0.3268	0.3536	0.3374	0.3416	0.2931	0.2851
	$t = 0.99$	0.3770	0.4067	0.3886	0.4034	0.3384	0.3309
Latin America and the Caribbean	MLD ($t = 1$)	0.6968	0.7212	0.7738	0.7112	0.6015	0.5829
	MLD between	0.0569	0.0525	0.0780	0.0747	0.0668	0.0627
	MLD within	0.6399	0.6687	0.6958	0.6365	0.5347	0.5202
	$t = 0.995$	1.3061	1.4891	1.6341	1.3721	1.0319	1.0021
	$t = 0.99$	2.0762	2.3767	2.6310	2.1094	1.5005	1.4298
Middle East and North Africa	MLD ($t = 1$)	0.6260	0.6570	0.5941	0.4720	0.5094	0.5890
	MLD between	0.1000	0.1091	0.0953	0.1065	0.1036	0.0521
	MLD within	0.5260	0.5479	0.4988	0.3655	0.4058	0.5369
	$t = 0.995$	1.0290	1.0846	0.9847	0.6869	0.7638	1.0557
	$t = 0.99$	1.3595	1.4341	1.3087	0.8620	0.9751	1.4424
South Asia	MLD ($t = 1$)	0.3982	0.4645	0.4701	0.5876	0.4256	0.4978
	MLD between	0.0267	0.0260	0.0216	0.0266	0.0263	0.0233
	MLD within	0.3715	0.4385	0.4486	0.5611	0.3993	0.4745
	$t = 0.995$	0.6120	0.8147	0.7999	0.8770	0.6809	0.7771
	$t = 0.99$	0.7870	1.1034	1.1218	1.2738	0.9077	1.0934
Sub-Saharan Africa	MLD ($t = 1$)	1.1867	1.1820	1.1800	1.0042	0.9565	0.9197
	MLD between	0.3218	0.3263	0.3447	0.3495	0.3335	0.3286
	MLD within	0.8648	0.8557	0.8353	0.6548	0.6230	0.5911
	$t = 0.995$	2.4769	2.1425	2.2123	1.6561	1.5469	1.3626
	$t = 0.99$	3.9883	3.1185	3.1917	2.4539	2.2747	1.9058
North America	MLD ($t = 1$)	0.2925	0.3315	0.2461	0.2454	0.2450	0.3869
	MLD between	0.0012	0.0017	0.0018	0.0018	0.0016	0.0018
	MLD within	0.2912	0.3297	0.2443	0.2436	0.2435	0.3851
	$t = 0.995$	0.3242	0.3854	0.3140	0.3060	0.2880	0.4779
	$t = 0.99$	0.3542	0.4362	0.3704	0.3579	0.3280	0.5598
Economies in transition	MLD ($t = 1$)	0.2324	0.5194	0.4809	0.5205	0.3569	0.3551
	MLD between	0.1279	0.1856	0.1922	0.1955	0.1482	0.1580
	MLD within	0.1045	0.3338	0.2887	0.3251	0.2087	0.1972
	$t = 0.995$	0.2442	0.6635	0.5912	0.6572	0.4107	0.4032
	$t = 0.99$	0.2551	0.8138	0.6915	0.7967	0.4586	0.4452

Note: The parameters of the GB2 distribution have been estimated by NLS using Eq. (4). MLD index in each country and year has been estimated by Monte Carlo simulation of synthetic samples of size $N = 10^6$.

Along with sub-Saharan Africa, Latin America and the Caribbean is the region where survey nonresponse resulted in the largest bias in the estimation of inequality levels. Without considering survey under-coverage, our estimates suggest that this region is also one of the most unequal regions. In 1990, Latin America presented an MLD of about 0.70, which rose to 0.77 by the end of the decade. Afterwards, inequality decreased for the rest of the period, which led to a reduction of 17 per cent in income inequality between 1990 and 2015. Income inequality estimates in South Asia are also strongly affected by survey under-coverage of top incomes. This is the only region that presents a steady increase in income inequality over the whole period. The Middle East and North Africa shows a decreasing trend during the 1990s that becomes positive afterwards until 2015. Interestingly, while inequality estimates without considering survey under-coverage of top incomes fell between 1990 and 2015, once the effect of non-response is imputed in our estimates, inequality in this region seems to increase over that period.

In North America, we find that inequality estimates that do not consider survey non-response remained relatively constant between 2000 and 2010, but the imputation of top income earners in the inequality levels leads to a decreasing trend during that period. This finding seems to confirm previous analyses on the effect of the financial crisis of 2008-09, which suggest that the richest top 1 per cent families experienced the largest loss in income immediately after the crisis, which in turn had a short-term ‘equalizing effect’ in the income distribution of the country (Alvaredo et al., 2013; Piketty and Saez, 2013). The declining trend in income inequality in the US rebounded after 2010, largely driven by a quick recovery in the growth of top incomes (Saez, 2015).

We observe similar inequality trends in Europe, which seem to mirror the North American patterns, except for the last spike in 2010. While in the US top income shares

fell sharply after 2008, in France, the United Kingdom and Germany the proportion of income accrued to the top decile did not show major variations (Piketty and Saez, 2013). As a result, European inequality remained constant between 2010 and 2015. A distinctive case is the one observed in South-Eastern Europe, the Commonwealth of Independent States and Georgia, which are classified by the UNDP as “economies in transition”.¹⁵ These group of countries initially exhibited low levels of income inequality. However, in a very short period of time, between 1990 and 1995, these countries saw such a sharp increase in income inequality that it was not offset by significant declines in inequality in more recent years, thus resulting in a staggering increase of 50 percent in income inequality between 1990 and 2015.

7 Conclusions

To date, the existing scholarship on economic inequality has been plagued by a lack of individual data. Most prior research on global inequality has approximated the world distribution of income using grouped data in the form of income shares. This approach introduces two potential sources of measurement error. Firstly, most of the existing works deploy nonparametric techniques, assuming equality of incomes within groups, which lead to downward-biased estimates of inequality. The second source of bias is caused by the under-coverage of top incomes in the household surveys from which income shares are generated. To develop reliable estimates of income inequality, we use a flexible parametric functional form to define plausible assumptions about the income distribution within shares. To consider the higher non-response rate of the richest individuals, our estimation strategy is designed to account for the truncated nature of the survey data

from which the income distribution of the whole population is obtained with a reliable representation of the upper tail.

Our results suggest that the bias of survey-based estimates might be substantively important. We find that, depending on the level of truncation, the size of the bias could be up to 40 per cent for inequality measures that are sensitive to the bottom of the income distribution. More importantly, we find that survey under-coverage of top incomes not only affects the levels of global inequality, but it could also change the direction of the trends. Disaggregating the analysis by world regions, we find that the effect of non-response among the richest individuals on inequality levels varies substantially across regions, with sub-Saharan Africa and Latin America being the most affected regions by the imputation of missing top incomes.

While we are confident about the robustness of our empirical strategy, there are aspects of the estimation of income inequality that we have not been able to fully incorporate within our research design. First, as for previous studies on income inequality, we were not able to accurately identify the non-response rates of household surveys. Harmonised tax and survey data are required to facilitate a fine-grained identification of the truncation point, which, unfortunately, are not periodically published. Hence, we set the non-response rate sufficiently small to ensure that it is below the actual non-response rate of all surveys, so our estimates are regarded as lower bounds.

Second, since income tends to be under-reported by the richest individuals, the results presented in this study may actually underestimate the “true” level of inequality. Recent methodological advances in the measurement of under-reporting require income micro-data on both surveys and tax records. Hence, due to the absence of periodical

tax records with comparable information to grouped survey data, the consideration of under-reporting issues remains limited.

Finally, prior research on income inequality has stressed the possibility that household surveys do not capture low incomes sufficiently well because the survey design does not target potential low-income respondents, or because low-income earners generally live in rural areas, thus being hard to contact them (Skoufias et al., 2001). Due to the lack of information about the proportion of the poor population not included in the survey, the consideration under-coverage issues at the bottom of the income distribution would add more uncertainty to our estimates on income inequality.

In sum, this study provides the most recent estimates of regional and global trends of income inequality. Although there are still issues to address in order to improve the measurement of income inequality, our method is a step forward towards improving our understanding of the impact of the richest on the evolution of global inequality and redistributive policy issues. As Atkinson et al. (2011) have rightly pointed out, more work is still needed to improve the coverage of household surveys and the accessibility to tax data for future research. In this sense, the methodological framework developed in this study is a powerful tool to mitigate the bias due to non-response of the very rich in household surveys, which might contribute to better monitoring of the progress towards the reduction of inequalities.

Notes

¹Realizing the Future we want for all. Report to the Secretary General prepared by the UN System Task Team to support the preparation of the Post-2015 UN Development Agenda, Draft V.1. April 2012: pp.11.

²A substantial body of research has also focused on national evolutions of top incomes share: see Saez, (2005) for the case of the US; Atkinson (2005) and Atkinson and Salverda (2005) for the UK; Piketty (2003) for France; Bach et al. (2013) for Germany; Roine and Waldenström (2008) for Sweden; Alvaredo and Londono-Velez (2013) for Colombia; Alvaredo (2009b) for Portugal; Atkinson et al. (2011), Andrews et al. (2011); Atkinson and Leigh (2008) for New Zealand.

³An alternative methodology that avoids defining ex-ante the shape of the distribution consist of estimating a non-parametric kernel distribution (Sala-i-Martin, 2006). While being a flexible model, its robustness has been questioned, particularly because of its poor performance at the tails (Dhongde and Minoiu, 2013).

⁴Previous studies have considered special or limited cases of this family, namely the Beta 2 distribution (Chotikapanich et al., 2012), the lognormal and the Weibull distributions (Pinkovskiy and Sala-i-Martin, 2014, and Chotikapanich et al., 1998, respectively) and the Lamé family (Jorda et al., 2014).

⁵Chen and Ravallion (2010) report that household surveys covered only 51.3 per cent of the world population in the early 1980s. By the mid-2000s, the coverage had increased to above 90 per cent.

⁶The problem of undersampling and underreporting of income at the upper tail of the income distribution are partly due to the way sampling frames are designed, but also due to attitudinal factors among the very rich. For a discussion, see Anand and Segal (2008).

⁷See McDonald (1984) and Kleiber and Kotz (2003) for details on the relation between the GB2 and its particular and limiting distributions.

⁸The WIID database is available on the following link: <https://www.wider.unu.edu/database/world-income-inequality-database-wiid34>

⁹For a review of the data coverage and the main statistical features of the WIID, see Jenkins (2015).

¹⁰Although the GB2 family is only able to represent one- and zero-mode distributions, Jorda et al. (2018) found that, even for bimodal distributions, lower bound estimates of inequality measures are not systematically more reliable than those provided by this parametric model.

¹¹ A detailed description of these guidelines can be obtained from <http://www.lisdatacenter.org/data-access/key-figures/methods/> and the R code used for the computation of inequality measures can be downloaded from <http://www.lisdatacenter.org/wp-content/uploads/files/access-key-programs-r-ineq.txt>

¹²Varying the sensitivity of inequality measures to the bottom or the upper tail is particularly relevant when there is no Lorenz dominance (Lambert, 2001).

¹³See the Appendix for more details on the regional classification of the countries.

¹⁴See Table A4 in the Appendix for regional inequality estimates based on the Theil index.

¹⁵See the Appendix for a full list of countries classified as economies in transition.

References

- Alvaredo, F. (2009a). The rich in Argentina over the twentieth century: 1932–2004. In Atkinson, A. and Piketti, T., editors, *Top incomes: A global perspective*. Oxford, Oxford University Press.
- Alvaredo, F. (2009b). Top incomes and earnings in Portugal 1936-2005. *Explorations in Economic History*, 46(4):404 – 417.
- Alvaredo, F. (2011). A note on the relationship between top income shares and the Gini coefficient. *Economics Letters*, 110(3):274–277.
- Alvaredo, F., Atkinson, A. B., Piketty, T., and Saez, E. (2013). The top 1 percent in international and historical perspective. Working Paper 19075, National Bureau of Economic Research.
- Alvaredo, F., Atkinson, A. B., Piketty, T., Saez, E., and Zucman, G. (2015). The world wealth and income database.
- Alvaredo, F. and Londoño, J. (2013). High incomes and personal taxation in a developing economy: Colombia 1993-2010. Commitment to Equity Working Paper, 12.
- Anand, S. and Segal, P. (2008). What do we know about global income inequality? *Journal of Economic Literature*, 46(1):57–94.
- Anand, S. and Segal, P. (2015). The global distribution of income. *Handbook of Income Distribution*, 2:937–979.
- Anand, S. and Segal, P. (2017). Who are the global top 1%? *World Development*, 95:111–126.

- Andrews, D., Jencks, C., and Leigh, A. (2011). Do rising top incomes lift all boats? *The BE Journal of Economic Analysis & Policy*, 11(1):1–45.
- Arnold, B. C. and Sarabia, J. M. (2018). *Majorization and the Lorenz order with applications in applied mathematics and economics*. Switzerland: Springer.
- Atkinson, A. B. (2005). Top incomes in the UK over the 20th century. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2):325–343.
- Atkinson, A. B. (2007). Measuring top incomes: Methodological issues. In Atkinson, A. and Piketty, T., editors, *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries*, pages 18–42. Oxford: Oxford University Press.
- Atkinson, A. B. (2017). Pareto and the upper tail of the income distribution in the uk: 1799 to the present. *Economica*, 84(334):129–156.
- Atkinson, A. B. and Brandolini, A. (2010). On analyzing the world distribution of income. *The World Bank Economic Review*, 24(1):1–37.
- Atkinson, A. B. and Leigh, A. (2008). Top incomes in New Zealand 1921–2005: Understanding the effects of marginal tax rates, migration threat, and the macroeconomy. *Review of Income and Wealth*, 54(2):149–165.
- Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71.
- Atkinson, A. B. and Salverda, W. (2005). Top incomes in the Netherlands and the United Kingdom over the 20th century. *Journal of the European Economic Association*, 3(4):883–913.

- Bach, S., Corneo, G., and Steiner, V. (2013). Effective taxation of top incomes in Germany. *German Economic Review*, 14(2):115–137.
- Behrens, K., Gaigné, C., Ottaviano, G. I., and Thisse, J.-F. (2007). Countries, regions and trade: On the welfare impacts of economic integration. *European Economic Review*, 51(5):1277–1301.
- Bhalla, S. S. (2002). *Imagine there’s no country: Poverty, inequality, and growth in the era of globalization*. Washington, DC: Peterson Institute.
- Blanchet, T., Fournier, J., and Piketty, T. (2017). Generalized pareto curves: theory and applications. WID.world Working Paper 2017/03.
- Bourguignon, F. and Morrisson, C. (2002). Inequality among world citizens: 1820-1992. *American Economic Review*, 92(4):727–744.
- Burkhauser, R. V., Héroult, N., Jenkins, S. P., and Wilkins, R. (2017). Top incomes and inequality in the UK: reconciling estimates from household survey and tax return data. *Oxford Economic Papers*, 70(2):301–326.
- Chen, S. and Ravallion, M. (2010). The developing world is poorer than we thought, but no less successful in the fight against poverty. *The Quarterly Journal of Economics*, 125(4):1577–1625.
- Chotikapanich, D., Griffiths, W. E., Hajargasht, G., Karunaratne, W., and Rao, D. (2018). Using the GB2 income distribution. *Econometrics*, 6(2):1–24.
- Chotikapanich, D., Griffiths, W. E., Prasada Rao, D., and Valencia, V. (2012). Global income distributions and inequality, 1993 and 2000: Incorporating country-level inequality modeled with beta distributions. *Review of Economics and Statistics*, 94(1):52–73.

- Chotikapanich, D., Valenzuela, R., and Rao, D. P. (1998). Global and regional inequality in the distribution of income: Estimation with limited and incomplete data. *Empirical Economics*, 22(4):533–546.
- Cowell, F. A. and Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics*, 141(2):1044–1072.
- Deaton, A. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and Statistics*, 87(1):1–19.
- Deininger, K. and Squire, L. (1996). A new data set measuring income inequality. *The World Bank Economic Review*, 10(3):565–591.
- Dhongde, S. and Minoiu, C. (2013). Global poverty estimates: A sensitivity analysis. *World Development*, 44:1–13.
- Diaz-Bazan, T. (2015). Measuring inequality from top to bottom. World Bank Policy Research Working Paper Series 7237.
- Dowrick, S. and Akmal, M. (2005). Contradictory trends in global income inequality: A tale of two biases. *Review of Income and Wealth*, 51(2):201–229.
- Hong, S., Han, H., and Kim, C. S. (2019). World distribution of income for 1970–2010: dramatic reduction in world income inequality during the 2000s. *Empirical Economics*. In press.
- Jenkins, S. P. (2009). Distributionally-sensitive inequality indices and the GB2 income distribution. *Review of Income and Wealth*, 55(2):392–398.
- Jenkins, S. P. (2015). World income inequality databases: an assessment of WIID and SWIID. *The Journal of Economic Inequality*, 13(4):629–671.

- Jenkins, S. P. (2017). Pareto models, top incomes and recent trends in uk income inequality. *Economica*, 84(334):261–289.
- Jorda, V., Sarabia, J. M., and Jäntti, M. (2018). Estimation of income inequality from grouped data. arXiv preprint arXiv:1808.09831.
- Jordá, V., Sarabia, J. M., and Prieto, F. (2014). On the estimation of the global income distribution using a parsimonious approach. *Research on Economic Inequality*, 22:115–145.
- Kakwani, N. (1980). On a class of poverty measures. *Econometrica: Journal of the Econometric Society*, 48(2):437–446.
- Kleiber, C. and Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. New Jersey and Canada: John Wiley & Sons.
- Krause, M. (2017). The millennium peak in club convergence: A new look at distributional changes in the wealth of nations. *Journal of Applied Econometrics*, 32(3):621–642.
- Lakner, C. and Milanovic, B. (2016). Global income distribution: From the fall of the berlin wall to the great recession. *World Bank Economic Review*, 30(2):203–232.
- Lambert, P. (2001). *The distribution and redistribution of income*. Manchester: Manchester University Press.
- Leigh, A. (2007). How closely do top income shares track other measures of inequality? *The Economic Journal*, 117(524):F619–F633.
- McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 52(3):647–665.

- McDonald, J. B. and Mantrala, A. (1995). The distribution of personal income: revisited. *Journal of applied Econometrics*, 10(2):201–204.
- McDonald, J. B. and Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, 66(1):133–152.
- Milanovic, B. (2011). *Worlds apart: Measuring international and global inequality*. Oxford:Princeton University Press.
- Monfort, P. and Nicolini, R. (2000). Regional convergence and international integration. *Journal of Urban Economics*, 48(2):286 – 306.
- Niño-Zarazúa, M., Roope, L., and Tarp, F. (2017). Global inequality: Relatively lower, absolutely higher. *Review of Income and Wealth*, 63(4):661–684.
- Piketty, T. (2003). Income inequality in France, 1901–1998. *Journal of Political Economy*, 111(5):1004–1042.
- Piketty, T. and Saez, E. (2013). Top incomes and the Great Recession: Recent evolutions and policy implications. *IMF Economic Review*, 61(3):456–478.
- Pinkovskiy, M. and Sala-i Martin, X. (2014). Africa is on time. *Journal of Economic Growth*, 19(3):311–338.
- Roine, J. and Waldenström, D. (2008). The evolution of top incomes in an egalitarian society: Sweden, 1903–2004. *Journal of Public Economics*, 92(1):366–387.
- Saez, E. (2005). Top incomes in the United States and canada over the twentieth century. *Journal of the European Economic Association*, 3(2-3):402–411.

Saez, E. (2015). Striking it richer: The evolution of top incomes in the united states (updated with 2012 preliminary estimates). Economics Department, UC Berkeley, Mimeo.

Sala-i Martin, X. (2006). The world distribution of income: falling poverty and... convergence, period. *The Quarterly Journal of Economics*, 121(2):351–397.

Skoufias, E., Davis, B., and De La Vega, S. (2001). Targeting the poor in mexico: an evaluation of the selection of households into progres. *World development*, 29(10):1769–1784.

UNU-WIDER (2017). World Income Inequality Database v.3.4.

8 Appendix

East Asia and the Pacific: American Samoa, Australia, Brunei Darussalam, China, Fiji, Micronesia, Fed. Sts., Guam, Hong Kong SAR, China, Indonesia, Japan, Cambodia, Kiribati, Korea, Rep., Lao PDR, Macao SAR, China, Marshall Islands, Myanmar, Mongolia, Northern Mariana Islands, Malaysia, New Caledonia, New Zealand, Philippines, Palau, Papua New Guinea, Korea, Dem. People Rep., French Polynesia, Singapore, Solomon Islands, Thailand, Timor-Leste, Tonga, Tuvalu, Vietnam, Vanuatu, Samoa.

Europe: Andorra, Austria, Belgium, Bulgaria, Switzerland, Channel Islands, Cyprus, Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, Faroe Islands, United Kingdom, Greece, Greenland, Croatia, Hungary, Ireland, Iceland, Italy, Liechtenstein, Lithuania, Luxembourg, Latvia, Monaco, Netherlands, Norway, Poland, Portugal, Romania, San Marino, Slovak Republic, Slovenia, Sweden, Turkey, Kosovo.

Latin America and the Caribbean: Aruba, Argentina, Antigua and Barbuda, Bahamas, The, Belize, Bolivia, Brazil, Barbados, Chile, Colombia, Costa Rica, Cuba, Curacao, Cayman Islands, Dominica, Dominican Republic, Ecuador, Grenada, Guatemala, Guyana, Honduras, Haiti, Jamaica, St. Kitts and Nevis, St. Lucia, St. Martin (French part), Mexico, Nicaragua, Panama, Peru, Puerto Rico, Paraguay, El Salvador, Suriname, Sint Maarten (Dutch part), Turks and Caicos Islands, Trinidad and Tobago, Uruguay, St. Vincent and the Grenadines, Venezuela, RB, Virgin Islands (U.S.).

Middle East and North Africa: United Arab Emirates, Bahrain, Djibouti, Algeria, Egypt, Arab Rep., Iran, Islamic Rep., Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Malta, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, Yemen.

South Asia: Afghanistan Bangladesh, Bhutan, India, Sri Lanka, Maldives, Nepal, Pakistan.

Sub-Saharan Africa: Angola, Burundi, Benin, Burkina Faso, Botswana, Central African Republic Cote d'Ivoire, Cameroon, Congo, Dem. Rep., Congo, Rep., Comoros, Cabo Verde, Eritrea, Ethiopia, Gabon, Ghana, Guinea, Gambia, The, Guinea-Bissau, Equatorial Guinea, Kenya, Liberia, Lesotho, Madagascar, Mali, Mozambique, Mauritania, Mauritius, Malawi, Namibia, Niger, Nigeria, Rwanda, Sudan, Senegal, Sierra Leone, Somalia, South Sudan, Sao Tome and Principe, Seychelles, Chad, Togo, Tanzania, Uganda, South Africa, Zambia, Zimbabwe.

North America: Bermuda, Canada, United States.

Economies in transition: Albania, Armenia, Azerbaijan, Bosnia and Herzegovina Belarus, Georgia, Kazakhstan, Kyrgyz Republic, Moldova, Macedonia, FYR, Montenegro, Russian Federation, Serbia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan.

Table A1: Summary statistics by region

	1990	1995	2000	2005	2010	2015
Number of Surveys	112	133	154	157	153	121
Years between the survey and the benchmark year (%)						
0	31	48	44	50	57	25
+/- 1	29	26	20	23	22	31
+/- 2	21	14	16	16	10	9
+/- 3	11	5	9	4	5	12
+/- 4	3	4	6	4	3	12
+/- 5	5	2	6	3	3	12
Income/ Consumption Sources (%)						
Income, disposable	37	41	38	34	40	50
Consumption	41	44	52	57	50	45
Income, gross	16	12	3	1	1	0
Others	6	3	8	8	9	5
Population covered (%)						
World	96	95	97	97	96	87
East Asia and the Pacific	96	96	96	96	97	87
Europe and Central Asia	94	99	100	100	100	100
Latin America and the Caribbean	96	96	97	97	96	92
Middle East and North Africa	76	76	83	73	72	46
South Asia	99	99	99	100	100	98
Sub-Saharan Africa	73	81	92	95	97	74
North America	100	100	100	100	100	100
Economies in transition	89	99	100	98	89	85
High income	95	97	98	97	95	82
Upper and middle income	96	98	99	97	97	92
Lower and middle income	94	95	96	97	97	89
Low income	64	70	86	91	91	76

Table A2: Regional income/consumption indices used to correct consumption shares (10 data points)

Region	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Developed	0.9382	1.0133	1.0284	1.0303	1.0270	1.0223	1.0260	1.0246	1.0074	0.9604
EAP	0.5566	0.6921	0.7464	0.7968	0.8399	0.8865	0.9303	0.9844	1.0591	1.2093
ECA	0.7791	0.8648	0.8940	0.9146	0.9337	0.9496	0.9661	0.9808	1.0035	1.1415
LAC	0.4461	0.6189	0.7040	0.7606	0.8078	0.8432	0.8819	0.9192	0.9811	1.2602
MENA	0.5493	0.7922	0.8346	0.8906	0.8928	0.8718	0.9211	0.8897	0.8967	1.2374
SA	0.5928	0.7275	0.7807	0.8214	0.8522	0.8857	0.9203	0.9586	1.0242	1.3011
SSA	0.4783	0.5717	0.6556	0.6897	0.7451	0.7802	0.8382	0.8861	1.0059	1.2609

Table A3: Regional income/consumption indices used to correct consumption shares (5 data points)

Region	Q1	Q2	Q3	Q4	Q5
Developed	0.9833	1.0294	1.0245	1.0252	0.9777
EAP	0.6342	0.7736	0.8650	0.9601	1.1590
ECA	0.8525	0.9307	0.9513	0.9803	1.0747
LAC	0.5530	0.7355	0.8273	0.9028	1.1730
MENA	0.7000	0.8651	0.8811	0.9033	1.1297
SA	0.6713	0.8025	0.8700	0.9410	1.2048
SSA	0.5350	0.6744	0.7643	0.8650	1.1860

Table A4: Regional income inequality and estimated bias due to survey under-coverage of top income shares. Theil index.

		1990	1995	2000	2005	2010	2015
East Asia and the Pacific	Theil index	0.9772	0.8036	0.7479	0.6212	0.4003	0.4368
	Theil between	0.7688	0.5664	0.4595	0.3231	0.1893	0.0754
	Theil within	0.2084	0.2372	0.2884	0.2981	0.2110	0.3613
	t = 0.995	1.0176	0.8594	0.7990	0.6793	0.4276	0.5032
	t = 0.99	1.0550	0.9103	0.8443	0.7315	0.4523	0.5660
Europe	Theil index	0.2347	0.2562	0.2418	0.2326	0.2183	0.2164
	Theil between	0.0707	0.0799	0.0761	0.0578	0.0461	0.0360
	Theil within	0.1641	0.1764	0.1656	0.1748	0.1722	0.1804
	t = 0.995	0.2568	0.2803	0.2650	0.2619	0.2441	0.2431
	t = 0.99	0.2749	0.2997	0.2838	0.2842	0.2642	0.2638
Latin America and the Caribbean	Theil index	0.5899	0.5891	0.6310	0.5883	0.5060	0.4764
	Theil between	0.0500	0.0459	0.0648	0.0603	0.0480	0.0452
	Theil within	0.5398	0.5432	0.5662	0.5280	0.4580	0.4312
	t = 0.995	0.7935	0.8237	0.8917	0.8031	0.6604	0.6294
	t = 0.99	0.9998	1.0505	1.1452	0.9992	0.7966	0.7567
Middle East and North Africa	Theil index	0.5118	0.5559	0.4906	0.4125	0.4445	0.4578
	Theil between	0.0965	0.1083	0.0931	0.1047	0.0988	0.0575
	Theil within	0.4153	0.4475	0.3975	0.3078	0.3457	0.4002
	t = 0.995	0.6579	0.7111	0.6277	0.4960	0.5464	0.6007
	t = 0.99	0.7626	0.8198	0.7229	0.5556	0.6182	0.6994
South Asia	Theil index	0.3340	0.3646	0.3791	0.5232	0.3483	0.4480
	Theil between	0.0295	0.0284	0.0226	0.0244	0.0233	0.0213
	Theil within	0.3044	0.3363	0.3565	0.4988	0.3251	0.4267
	t = 0.995	0.4314	0.5057	0.5026	0.6402	0.4563	0.5671
	t = 0.99	0.5013	0.6093	0.6040	0.7708	0.5384	0.6775
Sub-Saharan Africa	Theil index	1.2045	1.1590	1.1425	1.0603	0.9033	0.8604
	Theil between	0.3283	0.3288	0.3608	0.3464	0.3277	0.3237
	Theil within	0.8762	0.8302	0.7817	0.7139	0.5757	0.5367
	t = 0.995	1.5918	1.4715	1.4736	1.3473	1.1206	1.0085
	t = 0.99	2.0096	1.7479	1.7695	1.7249	1.3892	1.1630
North America	Theil index	0.3546	0.3846	0.2469	0.2536	0.2599	0.4393
	Theil between	0.0012	0.0017	0.0017	0.0017	0.0015	0.0017
	Theil within	0.3534	0.3829	0.2452	0.2519	0.2584	0.4376
	t = 0.995	0.3767	0.4183	0.2852	0.2888	0.2871	0.4896
	t = 0.99	0.3958	0.4471	0.3143	0.3164	0.3102	0.5300
Economies in transition	Theil index	0.1948	0.4548	0.4384	0.4687	0.3093	0.3073
	Theil between	0.0959	0.1439	0.1489	0.1456	0.1072	0.1150
	Theil within	0.0990	0.3109	0.2895	0.3230	0.2021	0.1923
	t = 0.995	0.2017	0.5050	0.4957	0.5393	0.3389	0.3339
	t = 0.99	0.2079	0.5484	0.5421	0.6033	0.3631	0.3557

Note: The parameters of the GB2 distribution have been estimated by NLS using Eq. (4). Theil index in each country and year has been estimated by Monte Carlo simulation of synthetic samples of size $N = 10^6$.