UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2016/94

Global inequality

How large is the effect of top incomes?

Vanesa Jorda[1] and Miguel Niño-Zarazúa[2]

August 2016

Abstract: In this paper, we estimate the recent evolution of global interpersonal inequality and examine the effect of omitted top incomes on the level and direction of global inequality. We propose a methodology to estimate the truncation point of household surveys by combining information on income shares from household surveys and top income shares from tax data. The methodology relies on a flexible parametric functional form that models the income distribution for each country-year point under different assumptions on the omitted information at the right tail of the distribution. Goodness-of-fit results show a robust performance of our model, supporting the reliability of our estimates. Overall, we find that the undersampling of the richest individuals in household surveys generate a downward bias in global inequality estimates that ranges between 15 per cent and 42 per cent, depending on the period of analysis, and the assumed level of truncation of the income distribution.

Keywords: inequality, top incomes, income distribution, truncated Lorenz curves

JEL classification: D31, E63, E01, O15

Acknowledgements: This study was written while Vanesa Jorda was at UNU-WIDER as a visiting scholar. The authors are grateful to participants at the UNU-WIDER internal seminar series for helpful comments on earlier versions of this paper. Vanesa Jorda wishes to acknowledge financial support from the Ministerio de Economía y Competitividad (project ECO2013-48326-C2-2-P).

[1] Department of Economics, University of Cantabria, Spain, corresponding author: jordav@unican.es; [2] UNU-WIDER, Helsinki, Finland.

This study has been prepared within the UNU-WIDER project 'World Inequality'.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-137-6

Typescript prepared by the authors.

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. Hence, redressing these trends will be a major challenge in the decades ahead.[1]

The rapid process of globalization has meant that labor and capital can move more easily across borders with important implications for global inequality trends. Empirical evidence suggests that, as a result of shifting industrial production from the North to the South, there has been an increase in inflows of capital into developing countries, which raised the demand for skilled workers in markets with abundant unskilled labor resources (Pavcnik, 2003). Consequently, the gap between wages has widened, thus pushing upwards within-country inequality trends, particularly in low-income countries (see Barro (2000), Kapstein and Milanovic (2002), Lundberg and Squire (2003) and also Kremer and Maskin (1996, 2006) for an alternative theoretical interpretation).

Free movement of capital has also meant that the richest can choose to reside in countries that offer lower marginal tax rates on income and capital, have higher living standards and also enjoy more advanced financial institutions that facilitate the reproduction of capital (Pritchett 1997). This is particularly relevant for the analysis of global inequality. Recent studies show that the highest income earners are significantly undersampled in household surveys (Alvaredo 2009a). Therefore, ignoring top incomes can generate substantial measurement errors and affect not only the levels, but also the trends of global inequality (Alvaredo 2011).[2]

There has been an increasing interest in literature on global inequality in the distributional dynamics of top incomes. This renewed interest has lead to important innovations both in data generation, notably the World Wealth and Income Database (WID) that includes series of top income shares from tax records (Alvaredo et al. 2015), and analytical methods that account for the bias generated from missing top incomes in the distributional analysis of a number of countries and world regions.[3] For the particular case of global inequality, Lakner and Milanovic (2013) and Anand and Segal (2016, 2015) have adopted methodologies that allow adjusting for the effect of under-reported top incomes in the estimation of global inequality. However, due

[1] 'Realizing 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

[2] Burkhauser et al. (2012) show that changes at the top 1% are responsible of most of the changes in the evolution of income inequality in the US.

[3] See Atkinson et al. (2011), Burkhauser et al. (2012), Piketty and Saez (2013), and Saez, (2005) for the case of the US; Atkinson (2005b) and Atkinson and Salverda (2005) for the UK; Piketty (2003) and Landais (2008) 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; Dell 2005 for Germany and Switzerland; Saez (2005) for the US and Canada; Atkinson et al. (2011), Andrews et al. (2011), Alvaredo et al. (2013) and Leigh (2007) for OECD countries; Atkinson and Leigh (2008) and Piketty and Saez (2013) for Anglo-Saxon countries.

to data limitations, they are forced to take an arbitrary threshold, the top 10 and 1 per cent respectively, as a truncation point.

The main contribution of this paper is to examine the effect of omitted top incomes on the level and the direction of global inequality. Most previous studies have used grouped data (quantile or decile distributions), while adopting a non-parametric approach with the implicit limitation of assuming equality of incomes within each income share (Niño-Zarazúa et al. 2016, Lakner and Milanovic 2013, Milanovic 2011, Bourguignon and Morrison 2002). These studies have reported, therefore, lower-bound estimates of global inequality.

To overcome this limitation, we propose an alternative method that consists of fitting a model to characterize the Lorenz curve and compute inequality measures from such a parametric model. 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 the household survey data. The estimated distribution would then provide us with corrected estimates for both average income and the level of inequality."

The parametric approach involves, however, the choice of a functional form that represents the distributional dynamics of income.[4] This is a challenging task because the analysis of global inequality involves a highly heterogeneous sample of countries in terms of income dynamics. To avoid misspecification bias, we use a well-suited functional form, the so-called "Generalized Beta of the second kind" (GB2) that nests the parametric assumptions in the literature (see McDonald 1984, Jenkins 2009). The GB2 is a general class of distributions that provides an accurate fit to income data (McDonald and Xu 1995, McDonald and Mantrala 1995). To our knowledge, this is the first study that adopts such a general model to fit the global income distribution.[5]

Our methodology allows us to consider the lower rate of response of the rich in our estimation, but in contrast to previous studies, we are able to define any level of truncation. This choice is arbitrary, so we resort to tax data available for a number of countries in the WID to determine the actual truncation point. We also present a sensitivity analysis on the assumptions on the truncation point to check the robustness of the results.

Overall, we find that the undersampling of the richest individuals in household surveys generate a downward bias in global inequality estimates that range between 15 and 42 per cent, depending on the period of analysis, and the assumed level of truncation of the income distribution. Our results also suggest that while the estimates of global inequality based on household surveys show a downward trend over the 1990-2010 period, the direction of such a trend can be reversed under specific assumptions about the level of under-sampling of the richest individuals.

[4] An 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 (Minoiu and Reddy 2007)

[5] Previous 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).

The remainder of the paper is structured as follows: Section 2 discusses measurement issues in the estimation of global inequality. Section 3 presents the adopted methodology. Section 4 describes the data used for the empirical analysis. The results are presented in Section 5, whereas Section 6 discusses the goodness-of-fit of our model. Finally, Section 7 concludes with some reflections on policy.

2 Measuring global income inequality

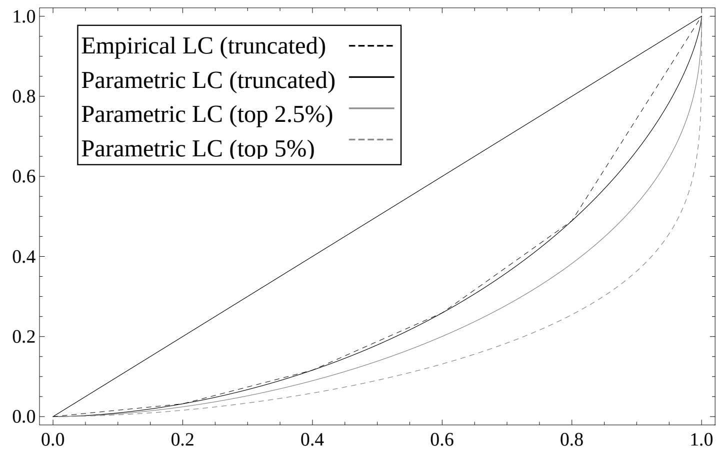
The concept of global income inequality is inherent to the disparities in income between all individuals in the world. In the utopia of having complete information about the income of all individuals in the world, or even a representative sample of the global population, the estimation of levels and trends of global inequality would be a relatively straightforward exercise. However, in the real world, there is not such a global survey of personal income, nor nationally presentative surveys for all countries in the world. Nevertheless, with the significant expansion in the generation of household surveys over the past 40 years, particularly in developing countries, there are now data available on per capita income (or consumption expenditure) for a significant number of countries over a reasonable long period of time.[6] This allow us to get a first approximation to the actual global income distribution under the assumption that all individuals in a country have the same income level, and to compute inequality measures for this hypothetical distribution. This is equivalent to obtain inequality measures using per capita income weighted by the population of each country, what Milanovic (2011) defines as 'Concept two inequality', and which we refer simply to as between-country inequality. This simple approach would yield a lower bound of the actual level of inequality since income disparities within countries would be suppressed. To provide more accurate estimates on global income inequality, additional information is needed on the within-country distributions.

Previous studies on global inequality have used income shares (typically five or ten points of the Lorenz curve) to estimate within-country inequality, to obtain measures of global interpersonal inequality (Bourgignon and Morrison 2002, Milanovic 2011, Lakner and Milanovic 2013, Niño-Zarazua et al. 2016, Dowrick and Akmal 2005). One limitation of these studies is the assumption that all individuals within each income share have the same income, thus suppressing inequality within each group. Hence, the results obtained by using this methodology can be regarded as being downward-biased estimates of the actual level of global inequality.

To illustrate this point graphically, we present in Figure 1 the Lorenz curve for the United States in 2013. The available information on within-country distributions are income shares, i.e. points of the Lorenz curve, represented by the black points in Figure 1. Computing inequality measures directly from income shares is a very intuitive strategy that avoids the need of imposing a model to fit the data. This approach implicitly assumes equality of incomes within income shares, which graphically implies linking these points linearly as illustrated by the dashed line in Figure

[6] Chen and Ravallion (2008) 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

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



Source: Authors' estimates.

1. This approach yields a Gini index of 0.442, while the estimate the actual Gini index using survey data is 0.477. We need, therefore, to define a model which allows us to impose more plausible assumptions on income dynamics within each income share to obtain reliable estimates of inequality. The solid black line in Figure 1 is an approximation to the Lorenz curve using an alternative parametric model. Despite the uncertainty about the within group distribution, the estimated Gini index of 0.473 yields a closer estimate to the actual survey value, which indicates that the model provides more reliable results.

A major drawback in using survey data in the estimation of within-country, and also global inequality, is the fact that the richest households respond survey questionnaires proportionately much less than the rest of the population, resulting in undersampling of top incomes. This explains why Alvaredo (2009a) could not find rich individuals reporting incomes over one million dollars in Argentina despite the fact there were about 700 people with such income levels according to tax records.[7] Therefore, by using household survey data, we are only able to estimate global inequality measures of the conditional distribution of income that is lower to a particular threshold b, i.e. $f(y|y < b)$.[8], whose Lorenz curve is represented by the solid black line in Figure 1.

In order to produce accurate estimates of global inequality, we need to model the income distribution of the entire population, which can be expressed more formally as follows:

$$f(y) = f(y|y < b)F(b), \tag{1}$$

[7]The 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).

[8]Albeit there is also left truncation due to the not consideration of the very poor, we do not consider it in our estimation. The effect of including them would be marginal compared with the effect of the very rich individuals and, due to the lack of information about the proportion of poor population not included in the survey, it would add more uncertainty to our estimates on income inequality.

where the probability $F(b)$ is the proportion of the population covered by the survey. We could obtain the income distribution from the conditional distribution in (1) quite straightforwardly if the threshold or the proportion of population covered by the survey were known, unfortunately, this information is unknown. This threshold plays a critical role in estimating the global distribution of income. The solid and dashed grey lines in Figure 1 represent the Lorenz curves of the unconditional distribution ($f(y)$) assuming that the survey covers 97.5 and 95 per cent of the population, respectively. The conditional distribution Lorenz dominates the unconditional distributions because top incomes are not included in the sample, thus being characterized by lower levels of inequality. It is quite intuitive to see that, as the proportion of 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 97.5% of the population, the Gini index of the unconditional distribution would be 0.568, whereas for a coverage rate of 95 per cent, it would rise to 0.692.

Lakner and Milanovic (2013) is the first study that quantified the effect of missing top incomes on global inequality estimates. In their methodology, the mean incomes from household surveys were adjusted to mean incomes from National Accounts Statistics (NAS), and the right tail reshaped according to a Pareto law at the top 10 per cent. More recently, Anand and Segal (2015, 2016) imputed the underresponse of the top 1 per cent of richest individuals from tax data while also adjusting the tail using the Pareto distribution. This semiparametric technique, albeit being an step forward in the analysis of global inequality, presents two limitations: Firstly, it is rather arbitrary to allocate all the excess of national accounts to survey means at the top decile. Whereas this choice was driven by data limitations, there is no empirical or theoretical justification to chose that particular threshold. Second, while the Pareto distribution is a valid candidate to model the right tail of the income distribution, previous studies have shown that this particular functional form performs very poorly for the bulk of the distribution. Due to the highly heterogeneous sample of countries that is involved in global analysis, the point where this model becomes suitable may differ across countries, thus causing misspecification bias.[9]

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. Our approach allows us to define any truncation point and address the existing variability of income within income shares. More specifically, we use a very general functional form, the so called Generalized Beta of Second Kind (GB2), which 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). This family nests most of the functional forms used to model income distributions, including the Beta 2 distribution used

[9]The Pareto law seems to fit adequately only the top 5 per cent of incomes in the United States and the United Kingdom (Dragulesku and Yakovenko, 2000).

by Chotikapanich et al. 2012, the lognormal and the Weibull distirbutions used by Pinkovskiy and Sala-i-Martin 2014 and Chotikapanich et al. 1998, respectively, and the Lamé family used by Jorda et al. 2014. This model is defined in terms of the probability density function (PDF) $(a, b, p, q \geq 0)$ as follows:

$$f(x; a, b, p, q) = \frac{ax^{ap-1}}{b^{ap}B(p, q)[1 + (x/b)^a]^{p+q}}, \quad x \geq 0, \tag{2}$$

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

We fit this model to all countries in each year. The use of the same functional form for the whole sample might not be a problem because the GB2 is a flexible model that includes a large number of parametric distributions as special cases. Therefore, it would converge to any of these sub-models if needed.[10]

As our estimation strategy relies on points of the Lorenz curve, we need to define it for the GB2 distribution. The Lorenz curve can be generally expressed as

$$L(u) = F_{Y_{(1)}}(F_Y^{-1}(u)), \tag{3}$$

where $F_{Y_{(1)}}(y) = \int_0^y tf(t)dt$ is the distribution of the first incomplete moment and $F_Y^{-1}(u)$ denotes the quantile function.

Thus, by substituting the explicit expressions of the quantile function and the first incomplete moment of the GB2[11] in Eq. (3), we get the Lorenz curve for this family:

$$L(u; a, p, q) = IB\left[\frac{[IB^{-1}(u; p, q)]^a}{1 + [IB^{-1}(u; p, q)]^a}; p + \frac{1}{a}, q - \frac{1}{a}\right], \tag{4}$$

where $IB(., .)$ stands for the incomplete beta function defined as $IB(y; p, q) = (1/B(p, q)) \int_0^y x^{p-1}(1-x)^{q-1}dx$.

If we estimate directly Eq.(4) using truncated survey data, which only include information about the bottom t% of the population, we would be estimating the parameters of the truncated distribution $(f(y|y < b))$. Hence the problem of undersampling top incomes would not be addressed. To do so, we would need to estimate the parameters of the distribution using a model that considers the right truncation of the data in the estimation. We use thus the following Lorenz curve to estimate the parameters of interest:

$$L(u|u < t) = \frac{L(u)}{L(t)}, \tag{5}$$

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 $(F(b))$ and $L(t)$ is the Lorenz curve at the truncation point,

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

[11]The expressions of the first incomplete moment and the quantile function can be found in Hajargasht et al. (2012).

i.e. the share of the total income held by the population covered in the survey. Substituting the formula of the truncated Lorenz curve of the GB2 in Eq. (5) we obtain,

$$L_t(u;a,p,q) = \frac{IB\left[\frac{[IB^{-1}(u;p,q)]^a}{1+[IB^{-1}(u;p,q)]^a}, p+\frac{1}{a}, q-\frac{1}{a}\right]}{IB\left[\frac{[IB^{-1}(t;p,q)]^a}{1+[IB^{-1}(t;p,q)]^a}, p+\frac{1}{a}, q-\frac{1}{a}\right]}. \tag{6}$$

The parameters of the distribution are estimated by minimizing the squared deviations between the income shares and the theoretical points of truncated Lorenz curve of the GB2 distribution in Eq.(6), that is

$$min_{a,p,q} \sum_{j=1}^{J} \left(\frac{IB\left[\frac{[IB^{-1}(u_j;p,q)]^a}{1+[IB^{-1}(u_j;p,q)]^a}, p+\frac{1}{a}, q-\frac{1}{a}\right]}{IB\left[\frac{[IB^{-1}(t;p,q)]^a}{1+[IB^{-1}(t;p,q)]^a}, p+\frac{1}{a}, q-\frac{1}{a}\right]} - s_j \right)^2. \tag{7}$$

The b parameter plays no role in the estimation because the Lorenz curve is independent to scale, so we use a moment estimator of the mean to estimate it. More specifically, we equal the theoretical expression for the mean of the GB2 distribution to the per capita income and solve for the b parameter:

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

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

Even when the parameters are estimated from a truncated Lorenz curve, we can obtain the Lorenz curve of the whole population by substituting them in Eq.(4). It is worth noting here that t is not a parameter to be estimated. It must be defined ex-ante, being the parameter estimates strongly affected by this choice. There is not a universal truncation point for all household surveys in the world; these are expected to differ across countries and years. Previous studies on top incomes have made different assumptions on the proportion of the population uncovered by household surveys, with 0.1, 1, 5 and 10 per cent being the most popular choices (Alvaredo 2011, Atkinson 2007, Lakner and Milanovic 2013). However, there is not any particular reason to set such particular thresholds.

As a tentative approximation to the truncation points, we resort to data on top incomes from the WID database (Alvaredo et al. 2015), and estimate the income distribution of the countries for which this information is available, to obtain the *optimal* truncation point. In particular, we define a grid for the t parameter from 0.90 to 0.999 by steps of 0.001, and for each truncation point we estimate Eq.(7).[12] We then estimate top income shares at 10, 5, 1, 0.1 per cent under our parametric model and compare them with those obtained from tax data from the WID. We

[12]Nonlinear regression techniques involve the definition of starting values for the optimization algorithm. The estimation of the GB2 distribution is characterized by multiple local minima, thus making it difficult to ensure that our estimates are those which globally minimize the squared sum of residuals (SSR). Hence, we consider a set of alternative initial values to ensure that our results are robust to the starting point of the optimization algorithm.

Table 1: Optimal truncation points for countries with comparable observations in WID and WIID

Country	year	average t	number of years
Australia	1986-2001	0.995	5
Canada	1975-2009	0.964	8
Finaland	1962-1987	0.994	2
Germany	1965-1977	0.991	2
Ireland	1987	0.995	1
Japan	1962-1987	0.988	27
Korea	1980-1998	0.985	6
Norway	1957-1963	0.986	2
The Netherlands	1962-1989	0.993	2
Singapore	1979	0.990	1
Sweden	1963-1988	0.981	10
Switzerland	1983	0.997	1
UK	1969-1999	0.992	8
US	1967-2004	0.976	36
Overall	1962-2009	0.983	111

Source: Authors' estimates using data from WIID and WID.

chose the truncation point which yields the lowest sum of the squared differences between both the estimated and the observed top incomes. This truncation point is optimal in the sense that it minimizes the squared differences between the top incomes available in WID and the estimated top incomes using the GB2 distribution.[13] Table 1 includes a summary of the truncation points obtained for the 111 country-years cases for which we have data on top incomes. Substantial differences are observed across countries, with the optimal truncation points ranging from 0.997 (Switzerland) to 0.964 (Canada). For this particular sample of countries, our results indicate that, on average, surveys represent 98.3 per cent of the population.

Regarding the evolution of the truncation point, our results suggest a decreasing trend in most of these countries, indicating that household surveys represent a lower proportion of the population over time.[14] This is partly due to the fact that, despite the very significant expansion in the generation of household surveys, particularly in developing countries, persistent problems of under-sampling and under-reporting of income at the upper tail of the income distribution remain due to the way sampling frames are designed and also to attitudinal factors among the richest,(Anand and Segal 2008). Notwithstanding, our sample is very limited, only including developed countries, so it would be arguably misleading to take the average truncation point from this subset of countries as a unique reference for the proportion of top incomes not included in

[13] We are aware of the fact that top income shares from tax data are obtained from left truncated samples because the poorest individuals are not required to pay taxes.Thus, the truncation point obtained by our procedure could be regarded as an upper bound of the actual truncation point

[14] Complete results on the truncation point for each country and year are available upon request.

household surveys around the world. This is because the rich in developing countries represent a much lower proportion of the population (Anand and Segal 2016). To provide complete estimates of the evolution of global inequality, we use different truncation points around this level to estimate Eq. (6) for each country and year.

Once the parameters of the distribution are estimated, the computation of inequality measures is relatively straightforward. The Lorenz curve is a powerful tool to measure inequality, but it leads to a partial ordering in the sense that not all distributions can be ranked. In these cases, we need inequality measures that provide a complete ordering of distributions. If two Lorenz curves cross, inequality measures can produce different results depending on their sensitivity to different parts of the distribution. The Gini coefficient is more sensitive to the middle of the distribution, and it does not allow us to change the weight given to differences in specific parts of the distribution. Varying the sensitivity of inequality measures to the bottom or the upper tail is particularly relevant when there is no Lorenz dominance (Lambert 2001).

To vary the importance of redistribution movements at different parts of the distribution, we compute an alternative set of inequality measures belonging to the generalized 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 give weights to the differences observed across the income distribution. To examine the evolution of global inequality, we compute GE measures for different parameter values. The mean log deviation (MLD) corresponds to the GE index 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 is equally sensitive to all parts of the distribution, being characterized by a parameter value equal to 1.[15]

The global income distribution can thus be defined as a mixture of the national distributions weighted by their population shares. To do so, let $Y_i, i = 1, ..., N$ be the income variable in county i, which is assumed to follow a GB2 distribution given by Eq.(2). Then, the global PDF can be expressed as,

$$f(y) = \sum_{i=1}^{N} \lambda_i f_i(y),$$

where λ_i stands for the population weights of country i.

Global inequality estimates of the GE measures can be derived by taking advantage of the decomposition of this family.

The especial cases given by the Theil and the MLD can be decomposed as follows:

$$T_W = \sum_{i=1}^{N} s_i T_i; T_B = \sum_{i=1}^{N} s_i log\left(\frac{\mu_i}{\mu}\right),$$

$$L_W = \sum_{i=1}^{N} \lambda_i L_i; L_B = \sum_{i=1}^{N} \lambda_i log\left(\frac{\mu}{\mu_i}\right),$$

[15] For the GB2 distribution, the closed expression for the GE measures can be found in Jenkins (2009).

where λ_i is the population share of the country i, s_i stands for the proportion of mean income of country i in the national mean: $s_i = \frac{\lambda_i \mu_i}{\mu} = \frac{\lambda_i \mu_i}{\sum_{i=1}^{N} \lambda_i \mu_i}$ and T_i and L_i are, respectively, the Theil index and the MDL of country i.

4 Data

For the analysis of global income inequality, we use data on income shares from UNU-WIDER's World Income inequality Database (WIID) version 3.3, which contains repeated cross-country information on Gini coefficients and income (or consumption) quantiles for 175 countries over the period 1960-2015.[16] The WIID is the most reliable and comprehensive database of worldwide distributional data currently available.[17] Our analysis focuses on the period 1990-2010 at five-year intervals – 1990, 1995, 2000, 2005 and 2010. As in each data point there were missing observations that reduce the coverage of the sample, we opted to include observations within a maximum of the previous/next five years of each data point, while preference was given naturally to the closest observations.

In addition, we adopt a conceptual base of the Camberra Group to minimize the problems that may arise from informational differences in the WIID in terms of unit of analysis, 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 included consumption data but adjusted by a correction procedure that partially harmonizes income and consumption data.

The methodology consist of comparing the average income shares with those of consumption, for the available country-year observations that had both income and consumption data available for the same year. If there were different sources for income and consumption data for a given country-year, our preference was to choose instances where both kinds of data came from the same sources. This was done in order to minimize measurement error due to variations in survey designs. We grouped countries into seven world regions and compute an average index of income relative to consumption (see Appendix Tables A2 and A1. Previous studies have used the absolute average difference between the two to correct either consumption shares (Niño-Zarazúa et al. 2016) or Gini indices (Deininger and Squire 1996). We opt, however, for the relative difference at regional, not global levels, to better account for the heterogeneity of countries in the income-consumption relationship. After these adjustments, we were able to cover over 90 per cent of

[16]The WIID database is available on the following link: https://www.wider.unu.edu/project/wiid-%E2%80%93-world-income-inequality-database

[17]For a review of the WIID, see Jenkins (2015).

global population in all year points (see Appendix Table A3).

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 2008, Milanovic 2011, Lakner and Milanovic 2013), most studies on global inequality have used national accounts, and in particular gross domestic product (GDP) per capita, given the limited availability of survey means (Atkinson and Brandolini 2010, Bhalla 2002, Bourguignon and Morrisson 2002, Dowrick and Akmal 2005, Niño-Zarazúa et al. 2016, 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 are likely to yield biased estimates of global inequality given the persistent undersampling and underreporting of income at the upper tail of the income distribution (Anand and Segal 2008).

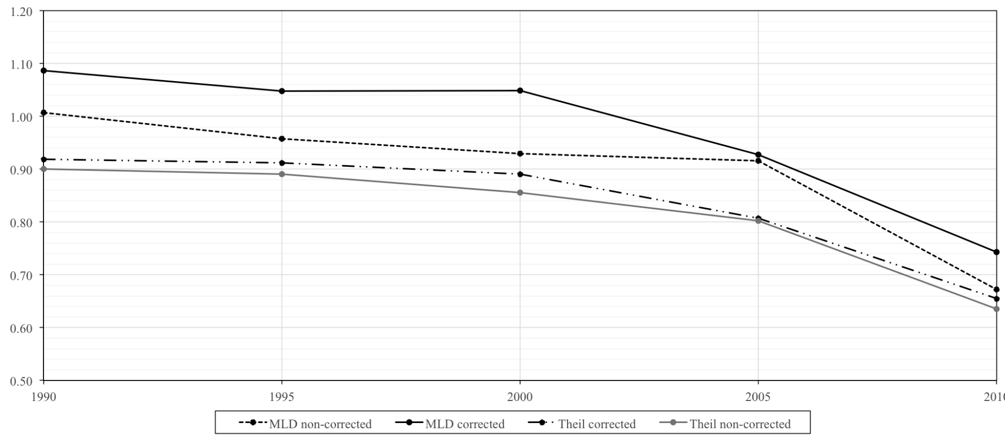
If discrepancies between national accounts and survey means were only driven by the undersampling and underreporting of top incomes, it would be reasonable to estimate the mean of the income distribution using data from national accounts. Deaton (2005) argues that GDP is a poor measure of household income 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. However, for the specific analysis of how top incomes may affect global inequality estimates, data from national accounts may actually be a better proxy for to the actual mean income of countries after accounting for the effect of the richest. For that reason, and also due to the limited coverage of survey means in the WIID, 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.

5 Results

In this section, we present an analysis of the evolution of global inequality for the period 1990-2010. We first estimate the levels of global inequality using group data without taking into account the inherent truncation of the data. These estimates can be regarded as a benchmark to assess the bias due to the omission of top incomes.

Before moving onto the analysis of global inequality, it is worth considering the effect of the correction of consumption shares in our estimates of global inequality. Figure 2 presents the evolution of the MLD and the Theil index before and after correcting for consumption shares. Both measures show higher levels of inequality when we apply the procedure described in Section 4 that makes income and consumption shares comparable. The results indicate that our procedure correctly reflects the expected relationship between income and consumption. Both series, corrected and non-corrected, show similar trends in both inequality measures, albeit some differing patterns in the period 2000-2010 are observed for the case of the MLD. We are certainly

Figure 2: The effect of correcting consumption shares on global inequality



Source: Authors' estimates based on data from WIID.

aware that, even when accounting for the heterogeneity across world regions, our procedure can correct, at least partially, for the divergence between these two welfare concepts.

Turning now to the evolution of global inequality, Table 2 presents the results of global inequality estimates based on the MLD and the Theil index. In line with previous studies, our estimates reveal a world characterized by extraordinarily high levels of income inequality, levels which are higher than the levels observed in the most unequal countries on earth. Such high levels of global inequality have nonetheless exhibited a declining trend over the past two decades, particularly in the 2000s. We also exploit the decomposability property by population subgroups of the GE measures to separate overall global inequality by its between- and within-country components. The decomposition analysis indicates that the decrease in global inequality has been mainly driven by a decline in between-country inequality, largely influenced by the rapid economic growth and convergence of income that populous countries such as China and India experienced over the past 30 years (Niño-Zarazua et al. 2016, Lakner and Milanovic 2013). In fact, the between-country inequality component accounts for two-thirds of the overall global inequality. Conversely, within-country inequality estimates show an increasing trend during the 1990s and early 2000s, but then a slightly decrease, which becomes particularly marked after the 2008 global financial crisis.

Thus far, we have conducted a conventional analysis of global inequality, without accounting for the effect of omitted top incomes. Based on the methodology proposed in Section 3, we present in Table 2 an estimation of the relative size of the bias of global inequality estimates, under different assumptions about the proportion of the population covered by household surveys. In Section 3, we conducted a parallel analysis using both survey data and tax records which revealed that, for a sample of developed countries, household surveys cover, on average, the bottom 98.3 per cent of the population. As pointed our earlier, the truncation points determined by our methodology are upper bounds, in the sense that tax data is left truncated and hence the top income shares of the whole society are expected to be higher. For that particular reason, we estimate truncation

Table 2: Global income inequality and estimated bias due to the omission of top income shares

		1990	1995	2000	2005	2010	1990-2000 Change (%)	2000-2010 Change (%)
	Total	1.0869	1.0458	1.0488	0.9275	0.7423	-3.5%	-11.4%
	Between	0.7399	0.6392	0.6013	0.5297	0.4115	-18.7%	-17.1%
	Within	0.3470	0.4066	0.4475	0.3978	0.3308	29.0%	-2.5%
MLD	Bias 0.995	-15.0%	-18.9%	-17.8%	-16.8%	-17.4%	-0.3%	-13.7%
	Bias 0.99	-24.1%	-28.6%	-27.2%	-29.5%	-27.8%	0.6%	-10.3%
	Bias 0.985	-31.4%	-35.9%	-36.6%	-36.0%	-32.7%	4.4%	-11.2%
	Bias 0.98	-38.7%	-40.8%	-41.9%	-42.3%	-38.1%	1.8%	-9.1%
	Total	0.9338	0.9341	0.9089	0.8138	0.6577	-3.1%	-11.5%
	Between	0.6915	0.6522	0.6276	0.5429	0.4118	-9.6%	-13.4%
	Within	0.2424	0.2819	0.2813	0.2709	0.2460	12.0%	-7.8%
Theil	Bias 0.995	-5.1%	-6.6%	-6.6%	-6.8%	-7.2%	-1.6%	-11.2%
	Bias 0.99	-8.8%	-11.1%	-11.0%	-11.8%	-12.3%	-0.7%	-10.8%
	Bias 0.985	-12.3%	-15.2%	-15.3%	-16.1%	-16.2%	0.2%	-10.6%
	Bias 0.98	-16.0%	-18.6%	-19.0%	-19.2%	-19.8%	0.5%	-10.8%

Source: Authors' estimates using data from WIID.

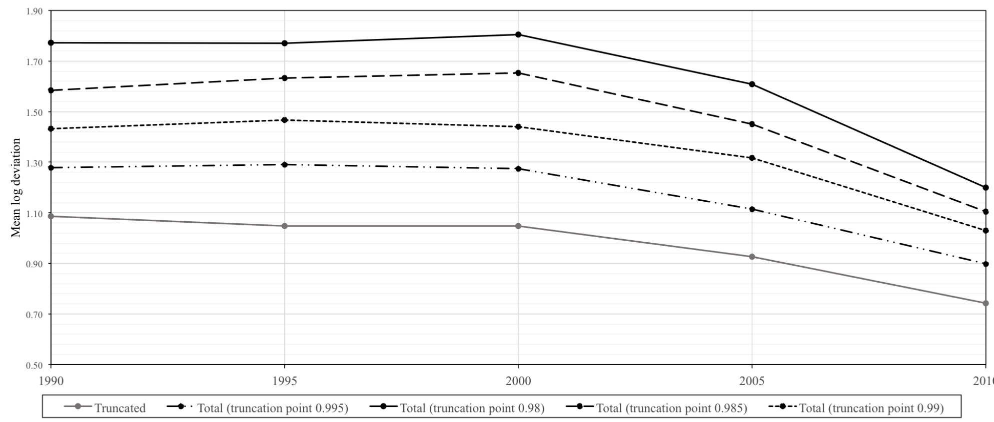
points at a 98 per cent levels. We note that our estimates are built upon the hypothesis that all countries have similar truncation points, an assumption that can be inconsistent with the heterogeneity observed in Table 1. Given the rigidity of this assumption,[18] our results can be interpreted as lower bounds of global income inequality in the sense that it is the minimum level of inequality that would exist if survey data represented that proportion of population or less in all countries.[19] Consequently, we opt for presenting global inequality estimates at relatively conservative rates of non-response among the rich: 0.5, 1, 1.5 and 2 per cent. We take such conservative approach even when we observe higher (actual) truncation points in some developed countries.

The results in Table 2 suggest that the bias in conventional global inequality estimates is substantially high even for the most conservative levels of truncation, which assume that household surveys represent 99.5 per cent of the population. Indeed, we find an underestimation of global inequality levels in the order of 15 and 19 percent using the MLD, and 5 and 7 per cent using the Theil index, depending on the year in question. The MLD shows a higher bias than the Theil index because it is more sensitive to the bottom of the distribution. As expected, the size

[18] Our methodology allows us to vary the truncation point across countries and years. However, due to the absence of information on the optimal truncation point for every country in the world, we are forced to assume similar truncation points for all countries.

[19] If some countries exhibited higher rates of non-response among the richest, than that assumed, their within-country inequality would be necessarily higher given that distributions with lower truncations points (i.e. with less proportion of population represented by the survey), Lorenz dominates the distributions with higher levels of truncation. This would imply that within-country inequality would be underestimated, thus characterizing our results as a lower bound of global inequality.

Figure 3: Global income inequality estimates based on MLD for different truncation points



Source: Authors' estimates based on data from WIID.

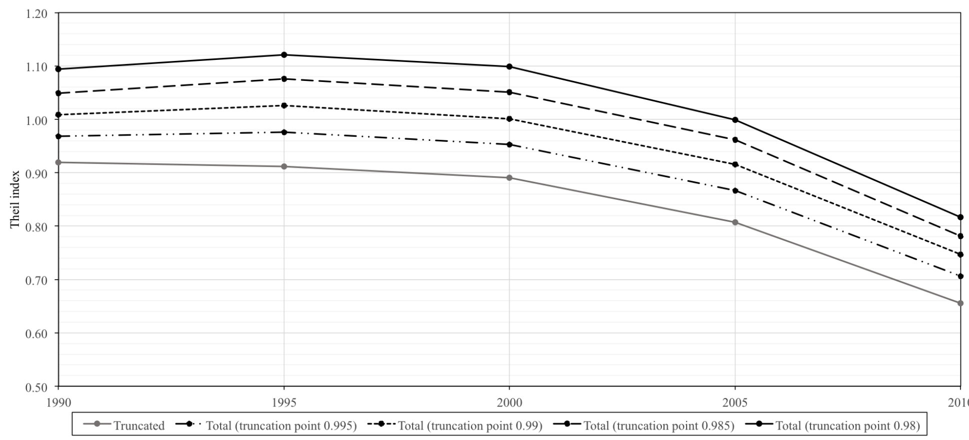
of the bias increases as the truncation point decreases, because distributions that represent a higher proportion of the population, Lorenz dominates those with lower truncation points. For a level of truncation similar to that observed in developed countries (98 per cent), global inequality estimates would be underestimated by 40 per cent using the MLD, and by 19 per cent using the Theil index.

The argument that previous estimates of global inequality are lower bounds has been repeatedly indicated in the literature. However, our results also show that omitted top incomes can not only change the level of global inequality, but also the direction of trends. This is illustrated in the last two columns of Table 1, which show the growth rate of inequality under different assumptions about the truncation level. The decreasing trend observed during the 1990s becomes positive when a truncation point of 99 per cent is considered for the MLD and 98.5 per cent for the Theil index.

In Figures 3 and 4 we illustrate graphically the recent evolution of global inequality adopting different assumptions about the proportion of omitted top incomes. Notwithstanding the decreasing trend in global inequality of the truncated sample, the MLD shows fairly different evolutions over the 1990s, depending on the level of truncation. Global inequality increased from 1990 to 1995 under all truncation points, except at 98 per cent, for which disparities remained constant. For the next five years, global inequality began to fall under the assumption of high representative coverage —at 99.5 and 99 per cent— of household surveys, whereas it increased if we assumed a high level of truncation. Similarly, the Theil index shows parallel trends of global inequality, except for the period 1990-95, which shows a modest reduction of global inequality, whereas an upward trend is observed once we impute omitted top incomes, no matter at what level of truncation.

The effect of top incomes can nonetheless vary across world regions, depending on the shape of the truncated distribution, and also the welfare institutions, fiscal policies and overall, the

Figure 4: Global income inequality estimates based on the Theil index for different truncation points



Source: Authors' estimates based on data from WIID.

social contracts that may dominate each of the world regions. In Table 3 we present inequality estimates based on the MLD for eight world regions based on the UNDP classification.[20] We compute the MLD assuming the following truncation points for the regional populations: 98.5, 99, 99.5 and 100 per cent. The results show that in contrast to what we observe at the global level, within-country inequality is the predominant driver of regional inequality. This indicates that world regions tend to be, on average, more homogenous in terms of per capita income. Only East Asia and the Pacific exhibited a higher between-country inequality in the 1990s although with a decreasing rate, relative to within-country inequality, during the 2000s. This is largely explained by market oriented structural reforms, technological change, trade liberalization 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 Fujita et al. (2001), Monfort and Nicolini (2000), Zeng and Zhao (2010), and also Krugman and Venables (1995) and Behrens et al. (2007) for a theoretical discussion). Furthermore, East Asia and the Pacific, together with the Middle East and North Africa are the world regions that have seen changes in the trends of inequality due to the imputation of top incomes. In East Asia and the Pacific, for instance, conventional inequality estimates, which suffer from undersampling bias of top incomes, show a reduction in inequality during the period 1990-95, but then an stagnation in inequality when we consider a truncation point at 99.5 per cent, and even an upward trend is observed when adopting higher truncation levels. Similar patterns are exhibited by South Asia, the Middle East and North Africa for the period 1995-2000.

In contrast, Europe, Latin America and the Caribbean and Transition Economies show parallel evolutions of regional inequality, with growing trends during the first half of the 1990s but then a decreasing one thereafter. This pattern is not sensitive to the inclusion of top incomes for

[20]See Table A4 in the Appendix for regional inequality estimates based on the Theil index.

any level of truncation. Interestingly, whereas North America displays an ascending trend in income inequality under conventional estimates for the period 2000-10, this trend is reversed in the period 2005-10 after adjusting for top incomes bias. Our findings confirm previous analyses on the effect of the financial crisis of 2008-09, particularly in the US, which indicate that the richest top 1 per cent families experienced the largest loss in income during, and 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).[21]

Our analysis has so far indicated that the effect of omitted top incomes on regional inequality is heterogeneous across regions, with a particularly marked effect in sub-Saharan Africa, the region with the lowest per capita income in the world. This suggests that the levels of inequality may be affected differently by the richest individuals, depending on the stage of development and income convergence of countries. In Table 4 we present the evolution of inequality under different assumptions on the truncation level of the distribution, based on the World Bank classification of countries by income groups.[22] Based on this classification, we observe that within-country inequalities drive predominately overall inequality in all income groups, with one exception: low-income countries. This is the only group whose between-country component dominates overall inequality, and also the only group that has not experienced convergence in income, but in fact divergence, during the period of analysis. Our results underpin the findings of earlier studies that show that in a era of rapid globalization, the growing adscription to South-South free trade agreements, and the resulting trade diversion instigated by changes in the comparative advantage of these nations, are key factors behind the absence of convergence between low-income countries (Baldwin et al. 2001, Venables 1999).

Turning our attention to the effect of omitted top incomes on overall inequality by income groups, we find a heterogeneous process, depending upon the level of income and also in some groups, the expected truncation point. For example, inequality in high-income countries increased from 1990 to 1995, remained constant in the second half of the 1990s and decreased steadily from 2000 to 2010. This trend is robust for any truncation point of the distribution. In contrast, for upper- and middle-income countries, inequality levels and trends vary considerably, depending on the selected truncation point. To illustrate, if we assume a truncation point of 99.5 per cent, inequality displays a decreasing trend from 1990 to 1995, while at higher levels of truncation, it increased over the same period. In contrast, for lower- and middle-income countries, inequality shows an upward trend between 1990 and 1995 assuming a truncation point of 99.5 per cent while it shows a downward trend at higher truncation levels. Thereafter, inequality remained constant until 2005, when it began to decrease. Finally, low-income countries experienced a reduction in inequality between 1990 and 2010 under relatively conservative assumptions about the truncation of the distribution, 100 per cent and 99.5 per cent, but reversed its trends between 2005 and 2010

[21] It is important to highlight the fact that the declining trend in income inequality in the US rebounded after 2010, largely driven by a quick recovery in growth of top incomes; see Saez (2015) for more details.

[22] Low-income countries are defined by the World Bank as those with a gross national income (GNI) per capita, calculated using the Atlas method, of US$1,025 or less in 2015; lower middle-income countries are those with a GNI per capita between $1,026 and $4,035; upper middle-income countries are those with a GNI per capita between $4,036 and $12,475; and high-income countries are those with a GNI per capita of $12,476 or more.

Table 3: Regional income inequality and estimated bias due to omitted top incomes

		1990	1995	2000	2005	2010
	MLD between	0.6359	0.4483	0.3446	0.2239	0.1374
	MLD within	0.2435	0.3422	0.4800	0.3392	0.2353
East Asia	MLD	0.8794	0.7905	0.8246	0.5631	0.3728
and the	MLD (Truncation 0.995)	0.9417	0.9415	0.9663	0.6580	0.4423
Pacific	MLD (Truncation 0.99)	1.0007	1.0923	1.0994	0.7621	0.5060
	MLD (Truncation 0.985)	1.0655	1.2230	1.3558	0.8466	0.5833
	MLD between	0.0799	0.0887	0.0812	0.0633	0.0510
	MLD within	0.2050	0.2093	0.1963	0.1927	0.1797
	MLD	0.2849	0.2981	0.2775	0.2560	0.2307
Europe	MLD (Truncation 0.995)	0.3404	0.3540	0.3239	0.3067	0.2717
	MLD (Truncation 0.99)	0.3918	0.4057	0.3639	0.3414	0.3044
	MLD (Truncation 0.985)	0.4430	0.4596	0.4069	0.3784	0.3403
	MLD between	0.0595	0.0528	0.0760	0.0734	0.0401
	MLD within	0.6503	0.6708	0.6443	0.5953	0.5358
	MLD	0.7098	0.7236	0.7203	0.6687	0.5758
Latin America	MLD (Truncation 0.995)	1.2260	1.3314	1.2808	1.0898	0.9182
and the	MLD (Truncation 0.99)	1.5311	1.6689	1.6351	1.4359	1.1726
Caribbean	MLD (Truncation 0.985)	1.9211	2.1814	2.0876	1.5903	1.1834
	MLD between	0.0843	0.0915	0.0842	0.0944	0.0909
	MLD within	0.3445	0.3435	0.3593	0.2596	0.2602
	MLD	0.4288	0.4350	0.4434	0.3540	0.3511
Middle East	MLD (Truncation 0.995)	0.5598	0.5589	0.5703	0.4284	0.4545
and North	MLD (Truncation 0.99)	0.6393	0.6172	0.6516	0.4909	0.4971
Africa	MLD (Truncation 0.985)	0.7255	0.6838	0.7227	0.5356	0.5435
	MLD between	0.0223	0.0208	0.0158	0.0220	0.0223
	MLD within	0.3578	0.4097	0.4198	0.5310	0.3941
	MLD	0.3801	0.4304	0.4356	0.5530	0.4164
South Asia	MLD (Truncation 0.995)	0.5528	0.6816	0.6526	0.7863	0.6469
	MLD (Truncation 0.99)	0.6706	0.8251	0.8885	1.2153	0.7904
	MLD (Truncation 0.985)	0.7487	0.9170	1.0744	1.4315	0.8157
	MLD between	0.3449	0.3531	0.3342	0.3303	0.3185
	MLD within	0.8650	0.7545	0.7038	0.4763	0.5348
	MLD	1.2099	1.1075	1.0380	0.8066	0.8533
Sub-Saharan	MLD (Truncation 0.995)	2.1655	1.8107	1.6811	1.2533	1.1560
Africa	MLD (Truncation 0.99)	3.0378	2.2936	1.9124	1.4459	1.5638
	MLD (Truncation 0.985)	3.8983	2.7929	2.2256	1.7064	1.8702
	MLD between	0.0013	0.0018	0.0018	0.0019	0.0016
	MLD within	0.2919	0.3333	0.2689	0.2902	0.2988
	MLD	0.2932	0.3352	0.2708	0.2920	0.3004
North	MLD (Truncation 0.995)	0.3253	0.3899	0.3383	0.3797	0.3640
America	MLD (Truncation 0.99)	0.3556	0.4442	0.4010	0.4615	0.4213
	MLD (Truncation 0.985)	0.3875	0.5003	0.4777	0.5473	0.4978
	MLD between	0.1194	0.1728	0.1836	0.1822	0.1369
	MLD within	0.0996	0.3705	0.2926	0.2220	0.2405
	MLD	0.2190	0.5433	0.4761	0.4043	0.3774
Economies	MLD (Truncation 0.995)	0.2278	0.7058	0.5820	0.4549	0.4509
in transition	MLD (Truncation 0.99)	0.2370	0.9012	0.6268	0.4987	0.5032
	MLD (Truncation 0.985)	0.2436	0.9319	0.6708	0.5369	0.5708

Source: Authors' estimates using data from WIID.

Table 4: Income inequality by development levels and estimated bias due to omitted top incomes

		1990	1995	2000	2005	2010
High income	MLD between	0.1885	0.2179	0.2155	0.1942	0.1407
	MLD within	0.2081	0.2544	0.2497	0.2282	0.2334
	MLD	0.3966	0.4723	0.4651	0.4224	0.3741
	MLD (Truncation 0.995)	0.4383	0.5354	0.5444	0.4835	0.4328
	MLD (Truncation 0.99)	0.4727	0.5852	0.5913	0.5352	0.4764
	MLD (Truncation 0.985)	0.5030	0.6352	0.6326	0.5866	0.5270
Upper and middle income	MLD between	0.4372	0.2522	0.1721	0.0989	0.0297
	MLD within	0.3216	0.4044	0.5332	0.4099	0.2750
	MLD	0.7589	0.6565	0.7053	0.5088	0.3047
	MLD (Truncation 0.995)	0.9036	0.8904	0.9350	0.7426	0.4174
	MLD (Truncation 0.99)	0.9986	1.0680	1.1340	0.8981	0.5643
	MLD (Truncation 0.985)	1.1361	1.3040	1.4422	1.0312	0.6242
Lower and middle income	MLD between	0.1287	0.1314	0.0992	0.0833	0.0736
	MLD within	0.4476	0.4744	0.4881	0.4830	0.4222
	MLD	0.5763	0.6057	0.5874	0.5663	0.4958
	MLD (Truncation 0.995)	0.9013	0.9288	0.8985	0.7751	0.7322
	MLD (Truncation 0.99)	1.1735	1.1503	1.1125	1.1164	0.8875
	MLD (Truncation 0.985)	1.4071	1.3351	1.3390	1.3109	0.9767
Low income	MLD between	0.5896	0.6168	0.5445	0.5999	0.5823
	MLD within	0.4909	0.6531	0.4453	0.3815	0.3847
	MLD	1.0805	1.2699	0.9898	0.9813	0.9669
	MLD (Truncation 0.995)	1.4089	1.8301	1.2295	1.1973	1.1716
	MLD (Truncation 0.99)	1.6964	2.2428	1.3517	1.3117	1.3379
	MLD (Truncation 0.985)	1.9557	2.3231	1.5663	1.3586	1.4929

Source: Authors' estimates using data from WIID.

Note: Theil index estimates by income groups are presented in Table A5 of the Appendix.

when considering higher levels of omitted top incomes.

6 Goodness-of-fit

How robust are our results? 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 therefore fundamental. We note though that due to the data structure of our sample, conventional tests of GOF cannot be performed.[23] We can, however, estimate the parameters of the GB2 distribution without considering the undersampling of top incomes. In particular, we use the income shares from household surveys available in the WIID to estimate (Eq.4). We then use these parameters to compute the Gini index, which we compare with the observed Gini index from household surveys. If the GB2 distribution is a good candidate to model income dynamics within income shares, the observed and the estimated Gini indices should be similar.

Our study focuses on the estimation of Eq.(6) to approximate the actual income distribution instead of the truncated one. However, it is not possible to assess the performance of the non-truncated version of the model because there is no available data to validate it. Despite this limitation, we argue that if the GB2 family performs well for truncated data, there is no reason to suspect that it would fail to do so for the entire distribution.

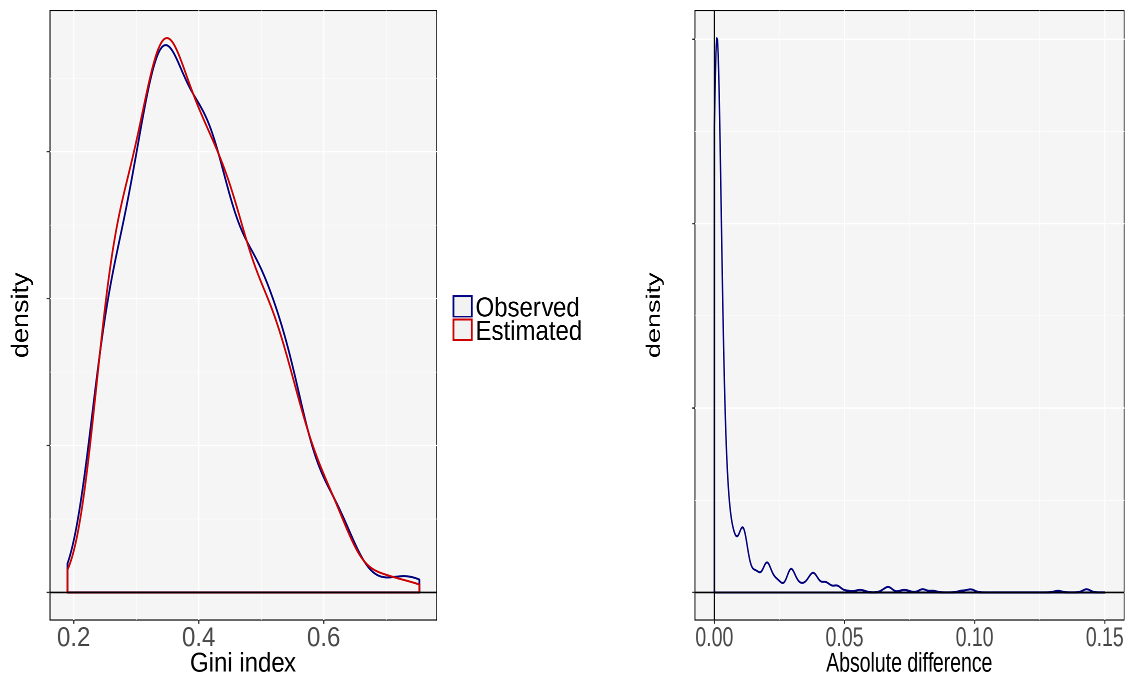
We estimate Eq.(4) for all country-year observations included in the sample. The left panel of Figure 4 shows the non-parametric density for the (observed) survey Gini index and the estimated Gini index using the GB2 distribution. The distribution of the survey Gini index and the estimated one is very similar, although we do not include nested observations in this figure. It might be possible that, while presenting very similar distributions, some particular countries exhibit a large gap between the observed and the estimated Gini index.

To analyse the deviance between the observed and the estimated Gini index for each country, we computed the absolute differences between these two, which are presented in the right panel of Figure 4. In 75% of the cases the gap between the survey Gini index and the estimate one is lower than 0.01 points. Absolute differences between 0.01 and 0.02 are found in 10% of the sample, whereas 11% of our estimates show divergences ranging between 0.02 and 0.05 points. Finally, about 3% of the estimates show differences greater than 0.05 points.[24] This confirms the robustness of our results and support previous studies that highlight the excellent performance of the GB2 distribution to model income data (McDonald 1984, Butler and McDonald 1986, Hajargasht et al. 2012).

[23] Bootstrapping techniques cannot be used in a context of limited information because we do not know the size of the sample, which has a strong impact on the bootstrap p-values.

[24] In some cases we find differences greater than 0.1: Kyrgyzstan (1996), Micronesia (2005), Singapore(2003), Zambia (2004).

Figure 4: Goodness-of-fit: density of the empirical and the estimated Gini index (left panel) and density of the absolute difference (right panel)



Source: Authors' estimates using data from WIID.

7 Conclusions

In this paper we have analysed the levels and evolution of income inequality over the period 1990-2010 with the aim of providing more accurate results than those found in previous studies. Most of the existing work on global inequality has used grouped data, and adopted empirical approaches, which implicitly assume equality of incomes within each income shares. Their results can be, therefore, characterized as lower bound estimates of global inequality. We overcome this limitation by using a flexible parametric functional form, the GB2 distribution, that allows us to estimate the income distribution for each country, and to model the differences in income within each income share. The GB2 includes, as particular cases, all the parametric functional forms that have been used in the past to model global income distribution, thus letting us to generate more reliable results with regard to trends and the levels of global inequality.

Another limitation of previous studies is the omission of top incomes in the estimation of global inequality. The resulting analysis from such truncated distributions at the upper tail can be regarded as providing lower bound (or downward biased) estimates of global inequality. In order to overcome this constraint, we propose a method to estimate the effect of undersampling of top incomes on the income distribution. Given the limited information about the truncation points of household surveys, which is expected to vary across countries, our method allows us to infer intuitively optimal truncation points based on tax data from a sample of countries, to correct for bias in global inequality estimates. Our results indicate that survey-based analyses of global inequality suffer from downward bias, which is in an order of magnitude of 15 and 42 per cent, depending on the period of analysis, and the assumed level of truncation at the upper tail of the income distribution. Furthermore, we find that omitted top incomes not only affect significantly the estimated levels of global inequality, but they can also change the direction of trends.

Disaggregating the analysis by world regions, we find that the effect of top incomes on the overall distribution varies significantly across regions, being sub-Saharan Africa, and the poorest countries in particular, the most affected by the imputation of top incomes on the levels and trends of inequality.

Given the relative small number of countries with available tax records, upon which our analysis has been based, and the uncertainly about the actual truncation points of the income distribution of countries, our results should be interpreted as the lowest level upon which global (or regional) inequality would lay if all countries were subject to a particular truncation level or higher in their income distribution due to undersampling problems in household surveys. As imperfect as it may be, our study is a step forward towards improving our understanding of the impact of the richest on the evolution of global inequality, and it also highlights the significance of our results for redistributive policy and also the titanic work that as Atkinson et al. (2011) have rightly pointed out, is still needed to improve the coverage and accessibility to tax data for future research.

Acknowledgements

This study was written while Vanesa Jorda was at UNU-WIDER as a visiting scholar. The authors are grateful to participants at the WIDER internal seminar series for helpful comments on earlier versions of this paper. Vanesa Jorda wishes to acknowledge financial support from the Ministerio de Economía y Competitividad (project ECO2013-48326-C2-2-P).

8 Appendix

Table A1: 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

Source: Authors' calculations based on data from WIID.

Table A2: 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

Source: Authors' calculations based on data from WIID.

Table A3: Summary statistics by region

	1990	1995	2000	2005	2010
Number of surveys	117	129	148	146	127
Years between the survey year and the benchmark year(%)					
0	26	46	43	47	52
+/- 1	28	28	21	22	20
+/- 2	26	13	14	18	9
+/- 3	9	7	9	5	7
+/- 4	5	3	6	5	5
+/- 5	5	3	7	1	7
Income/ Consumption Sources(%)					
Income	57	53	45	45	50
Consumption	31	30	23	12	18
Income/Consumption	12	16	32	42	32
Population covered(%)					
World	93	94	96	95	92
East Asia and the Pacific	96	96	99	96	96
Europe and Central Asia	94	95	96	96	96
Latin America and the Caribbean	96	96	97	97	95
Middle East and North Africa	75	75	87	77	68
South Asia	97	97	97	99	99
Sub-Saharan Africa	75	81	85	88	73
North America	100	100	100	100	100
Economies in transition	88	93	96	98	89
High income	97	97	98	94	91
Upper and middle income	95	96	97	96	94
Lower and middle income	93	94	97	98	95
Low income	69	71	77	81	68

Source: Authors' calculations based on data from WIID.

Table A4: Regional Theil index and estimated bias due to omitted top incomes

		1990	1995	2000	2005	2010
		1990	1995	2000	2005	2010
	Theil between	0.6359	0.4483	0.3446	0.2239	0.1374
	Theil within	0.2435	0.3422	0.4800	0.3392	0.2353
East Asia	Theil	0.8794	0.7905	0.8246	0.5631	0.3728
and the	Theil (Truncation 0.995)	0.9417	0.9415	0.9663	0.6580	0.4423
Pacific	Theil (Truncation 0.99)	1.0007	1.0923	1.0994	0.7621	0.5060
	Theil (Truncation 0.985)	1.0655	1.2230	1.3558	0.8466	0.5833
	Theil between	0.0799	0.0887	0.0812	0.0633	0.0510
	Theil within	0.2050	0.2093	0.1963	0.1927	0.1797
	Theil	0.2849	0.2981	0.2775	0.2560	0.2307
Europe	MLD (Truncation 0.995)	0.3404	0.3540	0.3239	0.3067	0.2717
	Theil (Truncation 0.99)	0.3918	0.4057	0.3639	0.3414	0.3044
	Theil (Truncation 0.985)	0.4430	0.4596	0.4069	0.3784	0.3403
	Theil between	0.0595	0.0528	0.0760	0.0734	0.0401
	Theil within	0.6503	0.6708	0.6443	0.5953	0.5358
	Theil	0.7098	0.7236	0.7203	0.6687	0.5758
Latin America	MLD (Truncation 0.995)	1.2260	1.3314	1.2808	1.0898	0.9182
and the	Theil (Truncation 0.99)	1.5311	1.6689	1.6351	1.4359	1.1726
Caribbean	Theil (Truncation 0.985)	1.9211	2.1814	2.0876	1.5903	1.1834
	Theil between	0.0843	0.0915	0.0842	0.0944	0.0909
	Theil within	0.3445	0.3435	0.3593	0.2596	0.2602
	Theil	0.4288	0.4350	0.4434	0.3540	0.3511
Middle East	Theil (Truncation 0.995)	0.5598	0.5589	0.5703	0.4284	0.4545
and North	Theil (Truncation 0.99)	0.6393	0.6172	0.6516	0.4909	0.4971
Africa	Theil (Truncation 0.985)	0.7255	0.6838	0.7227	0.5356	0.5435
	Theil between	0.0223	0.0208	0.0158	0.0220	0.0223
	Theil within	0.3578	0.4097	0.4198	0.5310	0.3941
	Theil	0.3801	0.4304	0.4356	0.5530	0.4164
South Asia	Theil (Truncation 0.995)	0.5528	0.6816	0.6526	0.7863	0.6469
	Theil (Truncation 0.99)	0.6706	0.8251	0.8885	1.2153	0.7904
	Theil (Truncation 0.985)	0.7487	0.9170	1.0744	1.4315	0.8157
	Theil between	0.3449	0.3531	0.3342	0.3303	0.3185
	Theil within	0.8650	0.7545	0.7038	0.4763	0.5348
	Theil	1.2099	1.1075	1.0380	0.8066	0.8533
Sub-Saharan	Theil (Truncation 0.995)	2.1655	1.8107	1.6811	1.2533	1.1560
Africa	Theil (Truncation 0.99)	3.0378	2.2936	1.9124	1.4459	1.5638
	Theil (Truncation 0.985)	3.8983	2.7929	2.2256	1.7064	1.8702
	Theil between	0.0013	0.0018	0.0018	0.0019	0.0016
	Theil within	0.2919	0.3333	0.2689	0.2902	0.2988
	Theil	0.2932	0.3352	0.2708	0.2920	0.3004
North	Theil (Truncation 0.995)	0.3253	0.3899	0.3383	0.3797	0.3640
America	Theil (Truncation 0.99)	0.3556	0.4442	0.4010	0.4615	0.4213
	Theil (Truncation 0.985)	0.3875	0.5003	0.4777	0.5473	0.4978
	Theil between	0.1194	0.1728	0.1836	0.1822	0.1369
	Theil within	0.0996	0.3705	0.2926	0.2220	0.2405
	Theil	0.2190	0.5433	0.4761	0.4043	0.3774
Economies	Theil (Truncation 0.995)	0.2278	0.7058	0.5820	0.4549	0.4509
in transition	Theil (Truncation 0.99)	0.2370	0.9012	0.6268	0.4987	0.5032
	Theil (Truncation 0.985)	0.2436	0.9319	0.6708	0.5369	0.5708

Source: Authors' estimates using data from WIID.

Table A5: Theil index by development levels and estimated bias due to omitted top incomes

		1990	1995	2000	2005	2010
	Theil between	0.1075	0.1354	0.1355	0.1218	0.0926
	Theil within	0.2217	0.2609	0.2297	0.2234	0.2352
	Theil	0.3292	0.3964	0.3653	0.3452	0.3277
High income	Theil (Truncation 0.995)	0.3498	0.4264	0.3985	0.3795	0.3593
	Theil (Truncation 0.99)	0.3665	0.4507	0.4233	0.4067	0.3837
	Theil (Truncation 0.985)	0.3831	0.4745	0.4484	0.4329	0.4097
	Theil between	0.4159	0.2555	0.1766	0.0995	0.0316
	Theil within	0.4211	0.4264	0.4950	0.3834	0.2643
	Theil	0.8370	0.6818	0.6716	0.4829	0.2960
Upper and middle income	Theil (Truncation 0.995)	0.9533	0.8158	0.7910	0.5889	0.3559
	Theil (Truncation 0.99)	1.0431	0.9198	0.8892	0.6620	0.4124
	Theil (Truncation 0.985)	1.1455	1.0447	0.9917	0.7410	0.4508
	Theil between	0.1526	0.1561	0.1168	0.0937	0.0794
	Theil within	0.3749	0.3826	0.3862	0.4157	0.3503
	Theil	0.5275	0.5387	0.5030	0.5094	0.4296
Lower and middle income	Theil (Truncation 0.995)	0.6535	0.6663	0.6239	0.5989	0.5302
	Theil (Truncation 0.99)	0.7536	0.7552	0.7100	0.7029	0.6007
	Theil (Truncation 0.985)	0.8421	0.8393	0.7955	0.7872	0.6548
	Theil between	0.4514	0.5076	0.4940	0.5570	0.5448
	Theil within	0.3204	0.5110	0.4029	0.2730	0.2664
	Theil	0.7719	1.0186	0.8970	0.8300	0.8112
Low income	Theil (Truncation 0.995)	0.8488	1.1599	0.9775	0.8890	0.8798
	Theil (Truncation 0.99)	0.9325	1.3034	1.0189	0.9429	0.9346
	Theil (Truncation 0.985)	1.0028	1.3095	1.1272	0.9574	0.9890

Source: Authors' estimates using data from WIID.

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 Gasparini, L. (2013). Recent trends in inequality and poverty in developing countries. Documentos de Trabajo del CEDLAS.

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. (2016). 'Who are the global top 1%?'. *International Development Institute Working Paper*, (2016-02).

Andrews, D., Jencks, C., and Leigh, A. (2011). 'Do rising top incomes lift all boats?'. *The BE Journal of Economic Analysis & Policy*, 11(1).

Arulampalam, W., Naylor, R. A., and Smith, J. P. (2004). 'A hazard model of the probability of medical school drop-out in the UK'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167(1):157–178.

Atkinson, A. B. (2005a). 'Comparing the distribution of top incomes across countries'. *Journal of the European Economic Association*, 3(2?3):393–401.

Atkinson, A. B. (2005b). '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: Mmethodological 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. and Brandolini, A. (2010). 'On analyzing the world distribution of income'. *The World Bank Economic Review*, page lhp020.

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.

Baldwin, R. E., Martin, P., and Ottaviano, G. I. (2001). 'Global income divergence, trade, and industrialization: The geography of growth take-offs'. *Journal of Economic Growth*, 6(1):5–37.

Barro, R. J. (2000). 'Inequality and growth in a panel of countries'. *Journal of Economic Growth*, 5(1):5–32.

Barro, R. J. (2012). 'convergence and modernization revisited'. Technical report, National Bureau of Economic Research.

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.

Bourguignon, F. and Morrisson, C. (2002). 'Inequality among world citizens: 1820-1992'. *American Economic Review*, 92(4):727–744.

Burkhauser, R. V., Feng, S., Jenkins, S. P., and Larrimore, J. (2012). 'Recent trends in top income shares in the United States: reconciling estimates from march CPS and IRS tax return data'. *Review of Economics and Statistics*, 94(2):371–388.

Butler, R. J. and McDonald, J. B. (1986). 'Trends in unemployment duration data'. *The Review of Economics and Statistics*, pages 545–557.

Chen, S. and Ravallion, M. (2008). The developing world is poorer than we thought, but no less successful in the fight against poverty. World Bank Policy Research Working Paper Series, 4703.

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.

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.

Dell, F. (2005). 'Top incomes in Germany and Switzerland over the twentieth century'. *Journal of the European Economic Association*, 3(23):412–421.

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.

Dragulescu, A. and Yakovenko, V. M. (2000). 'Statistical mechanics of money'. *The European Physical Journal B-Condensed Matter and Complex Systems*, 17(4):723–729.

Fujita, M., Krugman, P. R., and Venables, A. (2001). *The spatial economy: Cities, regions, and international trade*. Cambridge, MA: MIT press.

Gomes, O., Combes, C., and Dussauchoy, A. (2008). 'Parameter estimation of the generalized gamma distribution'. *Mathematics and Computers in Simulation*, 79(4):955–963.

Hajargasht, G., Griffiths, W. E., Brice, J., Rao, D. P., and Chotikapanich, D. (2012). 'Inference for income distributions using grouped data'. *Journal of Business & Economic Statistics*, 30(4):563–575.

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.

Jordá, V. and Sarabia, J. M. (2014). 'Well-being distribution in the globalization era: 30 years of convergence'. *Applied Research in Quality of Life*, pages 1–18.

Jordá, V., Sarabia, J. M., and Prieto, F. (2014). 'On the estimation of the global income distribution using a parsimonious approach'. In *Economic Well-Being and Inequality: Papers from the Fifth ECINEQ Meeting (Research on Economic Inequality)*, volume 22, pages 115–145.

Kakwani, N. (1980). 'On a class of poverty measures'. *Econometrica: Journal of the Econometric Society*, pages 437–446.

Kapstein, E. B. and Milanovic, B. (2002). *When markets fail: Social policy and economic reform*. New York: Russell Sage Foundation.

Kleiber, C. and Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. New Jersey and Canada: John Wiley & Sons.

Kremer, M. and Maskin, E. (1996). Wage inequality and segregation by skill. Discussion Paper of the National Bureau of Economic Research, 5718.

Kremer, M. and Maskin, E. (2006). Globalization and inequality. Mimeo.

Krugman, P. and Venables, A. J. (1995). 'Globalization and the inequality of nations'. *Quarterly Journal of Economics*, 110(4):857–880.

Lakner, C. and Milanovic, B. (2013). Global income distribution: From the fall of the berlin wall to the great recession. World Bank Policy Research Working Paper, 6719.

Lambert, P. (2001). *The distribution and redistribution of income*. Manchester University Press.

Landais, C. (2008). Top incomes in France: booming inequalities? Paris School of Economics, Mimeo.

Leigh, A. (2007). 'How closely do top income shares track other measures of inequality?'. *The Economic Journal*, 117(524):F619–F633.

Lundberg, M. and Squire, L. (2003). 'The simultaneous evolution of growth and inequality'. *The Economic Journal*, 87(1):1–19.

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*. Princeton University Press.

Minoiu, C. and Reddy, S. G. (2007). Kernel density estimation based on grouped data: The case of poverty assessment. Available at SSRN 991503.

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. (2016). 'Global inequality: Rlelatively lower, absolutely higher'. *Review of Income and Wealth*.

Parker, S. C. (1999). 'The generalised beta as a model for the distribution of earnings'. *Economics Letters*, 62(2):197–200.

Pavcnik, N. (2003). 'What explains skill upgrading in less developed countries?'. *Journal of Development Economics*, 71(2):311–328.

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.

Pritchett, L. (1997). 'Divergence, big time'. *The Journal of Economic Perspectives*, 11(3):3–17.

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.

Sarabia, J. M. and Jordá, V. (2014). 'Explicit expressions of the Pietra index for the generalized function for the size distribution of income'. *Physica A: Statistical Mechanics and its Applications*.

Unesco (2014). Statistical yearbook. Paris.

UnitedNations (2014). *The Millennium Development Goals Report 2008*. United Nations Publications.

Venables, A. J. (1999). Regional integration agreements: a force for convergence or divergence? World Bank Policy Research Working Paper,2260.

Wail, B., Hanchane, S., and Kamal, A. (2012). 'Educational inequality in the world, 1950–2010: Estimates from a new dataset'. *Research on Economic Inequality*, 20:337–366.

Zeng, D.-Z. and Zhao, L. (2010). 'Globalization, interregional and international inequalities'. *Journal of Urban Economics*, 67(3):352–361.