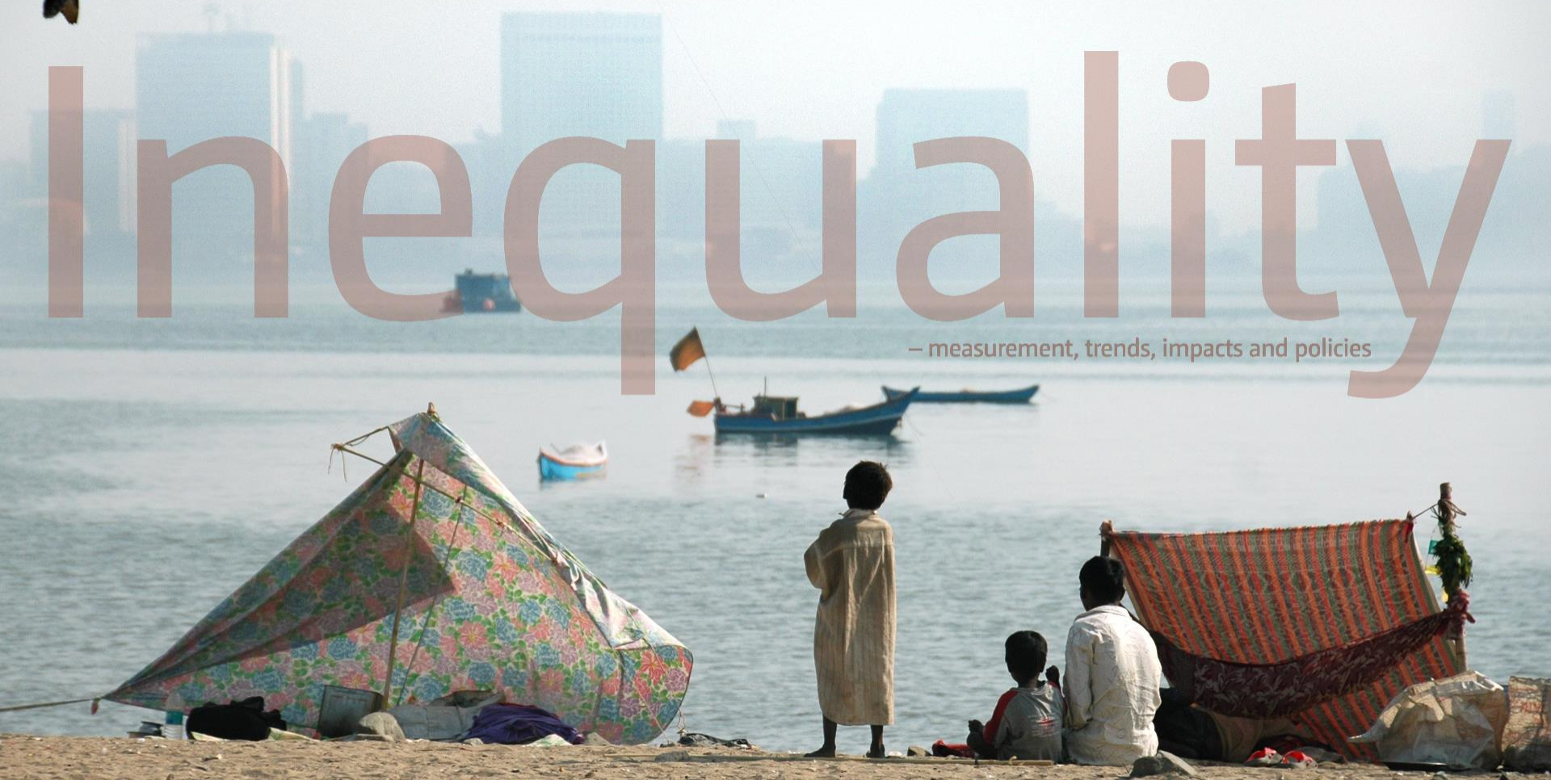


Income inequality: What is the effect of omitted top incomes?

Miguel Niño-Zarazúa, UNU-WIDER
Vanesa Jorda, University of Cantabria

13th AFD Conference on Inequality and Social Cohesion, Paris 2018

UNITED NATIONS UNIVERSITY
UNU-WIDER

Outline of presentation

1. The concern of **income inequality** with a focus on the recent evolution of global and regional trends

2. Data issues

   - Dealing with omitted top incomes in the estimation of (global, regional also within-country) inequality

   - Parametric methods to correct the bias arising from the truncation of the income distribution from household surveys

2. Global and regional inequality estimates

3. Concluding remarks

Results in a nutshell

- In this paper, we propose a methodology to estimate the effect of omitted top incomes on the levels and the evolution of income inequality over the period 1990-2010

- To do so, we fit a model to characterize the Lorenz curve and compute inequality measures from a parametric model based on the so-called "Generalized Beta of the second kind" (GB2) distribution

- Our results indicate that survey-based analyses of global inequality suffer from downward bias in the 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

- The proposed methodology is a step forward towards improving our understanding of the impact of the rich on the evolution of global and regional inequality and **highlights the importance of improving the coverage and accessibility to tax data for better policy design**

The concern of income inequality

- Inequality plays a critical role in the success of development trajectories of countries
- High inequality reduces the efficacy of economic growth to poverty reduction (Ravallion 2001)
- Inequality also affects a country's potential of economic growth, by impacting negatively on **human capital formation**
- Negative implications of high levels of inequality, in terms of **social cohesion and crime**, (Kelly, 2000), **conflict and political instability** (Alesina and Perotti, 1996) and **corruption and governance** (You and Khagram, 2005) are widely acknowledged

The concern of income inequality

- The rapid process of globalization in past decades has meant that **labor and capital can move more easily across borders**

- Shifting industrial production from the North to South has led to increases in the inflows of capital into developing countries, which raised the demand for skilled workers in markets with abundant unskilled labor resources (Pavcnik, 2003)

- The gap between wages has widened, thus pushing upwards within-country inequality trends globally (Niño-Zarazúa et al 2016, Barro 2000, Kapstein and Milanovic 2002, Lundberg and Squire 2003)

- Free movement of capital has also meant that the rich can choose to live (or keep their assets) 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

The concern of income inequality

- This is particularly relevant for the analysis of global, regional and domestic inequality. Recent studies show that the **highest income earners are significantly undersampled in household surveys** (Alvaredo 2009)

- Ignoring top incomes can generate substantial measurement errors and affect not only the levels, but also the trends of income inequality

- There have been important innovations both in: i) **data generation across countries** (e.g. World Wealth and Income Database) that includes series of top income shares from tax records (Alvaredo et al. 2015), and ii) **analytical methods** that account for the bias arising from missing top incomes in the estimation of income inequality

- Unfortunately, **tax data remains not openly accessible for most countries**

Data issues

What can be done in the absence of tax records, besides making tax authorities more transparent and accountable for their citizens?

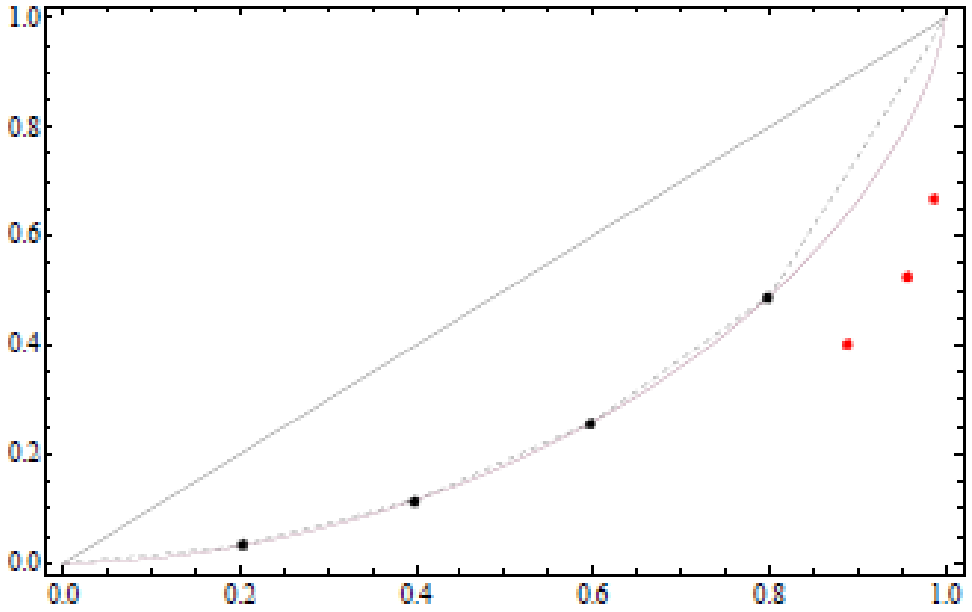
Dealing with omitted top incomes

- In this paper, we propose a methodology that consists of fitting a model to characterize the Lorenz curve and compute inequality measures from a parametric model based on the so-called "Generalized Beta of the second kind" (GB2) distribution.

- The GB2 is a general class of distributions that provides an accurate fit to income data (McDonald and Xu 1995, McDonald and Mantrala 1995), and nests the parametric assumptions in the literature (see McDonald 1984, Jenkins 2009).

- Our methodology allows us to consider the lower rate of response of the rich in HH surveys (truncation points), by using both **income shares from household surveys** (available in the UNU-WIDER's WIID database) and **top income shares from tax data from WID**

Dealing with omitted top incomes



Underestimation: Using tax data we observe that the LC is much farther away from the egalitarian line!

WHY? TRUNCATION

Dealing with omitted top incomes

MAIN CHALLENGE: Use truncated information on income shares to obtain reliable estimates of the whole income distribution.

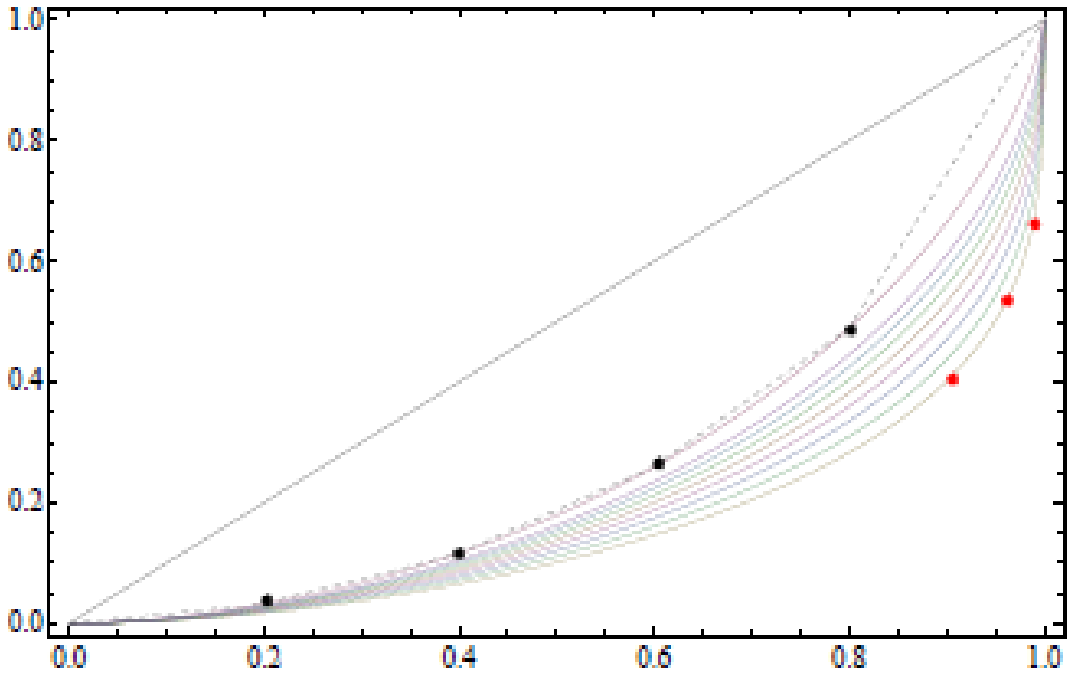
The truncated LC ($L^*(u)$) and the whole LC ($L(u)$) are related as follows:

$$L^*(p) = \frac{L(p)}{s_t} \rightarrow L^*(p) = \frac{L(p)}{L(t)} \rightarrow L(p) = L^*(p)L(t)$$

where s_t is the income share of the population represented in the survey, which can be expressed as the value of the LC at the truncation point t

PROBLEM: The truncation point t is unknown, and hence $L(t)$ cannot be directly computed.

Dealing with omitted top incomes



We chose the truncation point which yields the "closest" LC to the top income shares from tax data

Dealing with omitted top incomes

- The main assumption of this fully parametric methodology is that the model is fitted to all countries over the whole period.
- **Since the parameters of the GB2 vary across countries and over time, it allowing us to model the potential changes in the income distribution**

The Generalized Beta distribution of the Second Kind (GB2)

Let X be a non-negative random variable representing the individual income, which follows a GB2 distribution with probability density function (PDF) given by,

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

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 b is a scale parameter.

Why GB2?

1. General class of distributions that is acknowledged to provide an accurate fit to income data (McDonald and Xu, 1995; McDonald and Mantrala, 1995).
2. Includes all the distributions used to model the global distribution of income:
   - The beta 2 distribution used in (Chotikapanich et al., 2012)
   - The lognormal and the Weibull distributions (Pinkovskiy and Sala-i-Martin; Chotikapanich et al., 1997)
   - The Lamé distributions (Particular cases of Dagum and Sigh-Maddala distributions) (Jorda et al., 2013)
3. It is a good candidate to model censored data (Jenkins et al., 2011; Burkhauser et al., 2011)

The model

The Lorenz curve can be generally expressed as,

$$L(u) = F_{X_{(1)}}(F_X^{-1}(u)),$$

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

The Lorenz curve of the GB2 distribution

Let X be a non-negative random variable which follows a GB2 distribution, the Lorenz curve of X is given by,

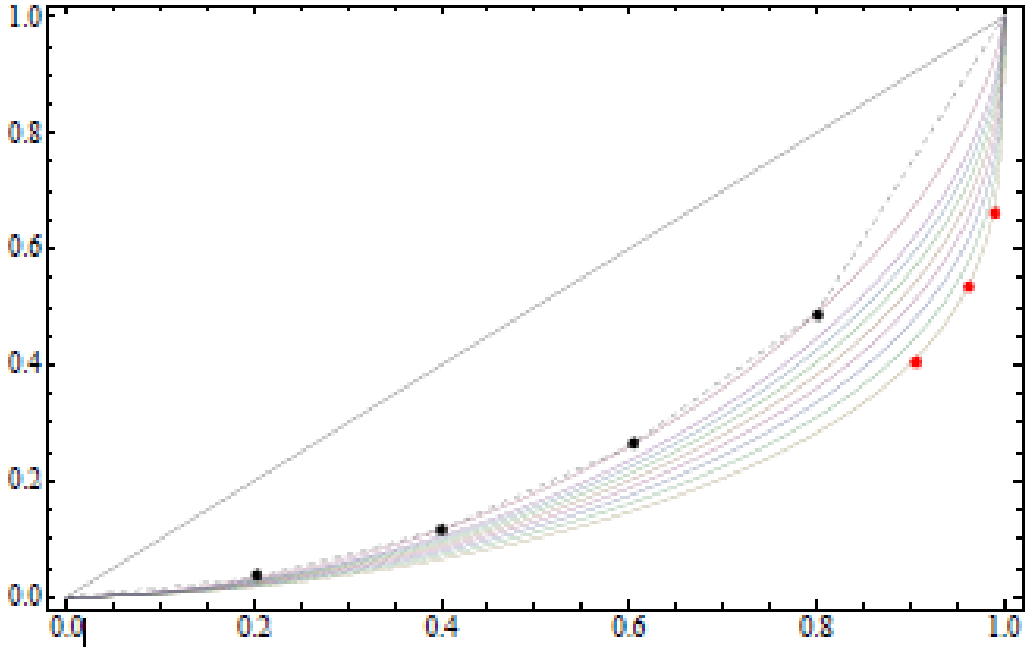
$$L(u; a, p, q) = IB\left[\frac{[IB^{-1}(u; p, q)]^a}{1 + [IB^{-1}(u; p, q)]^a}; p + \frac{1}{a}, q - \frac{1}{a}\right],$$

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

Estimation methods

- It is worth noting that **t** must be defined ex-ante because it is not a parameter to be estimated
- Previous studies on top incomes have made different assumptions on the proportion of the population uncovered by household surveys, being 0.1, 1, 5 and 10 per cent the most popular choices (Alvaredo 2011, Atkinson 2007, Lakner and Milanovic 2013)
- Setting an *arbitrary* threshold may bias the results because the estimation of the income distribution is strongly affected by this choice
- To minimize the bias due to discretionary thresholds, **we approximate the actual truncation points by resorting to data on top incomes from tax records from a sample of countries**

Optimal truncation point



We chose the truncation point which yields the minimum difference between the empirical top income shares (red points) and the estimated income shares.

Optimal truncation point

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

Inequality measures

Generalized Entropy measures

$$I(\theta) = \sum_{i=1}^{N} \lambda_i^{1-\theta} s_i^{\theta} I_i^{(\theta)} + \frac{1}{\theta(\theta-1)} \left(\sum_{i=1}^{N} \lambda_i^{1-\theta} \left(\frac{\mu_i}{\mu} \right)^{\theta} - 1 \right).$$

λ_i and $I_i^{(\theta)}$ are the population share and the GE measure of the country i. s_i stands for the proportion of income of the country i in the global mean income.

Theil's Entropy index

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

Mean Log Deviation

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

Estimating the (global) income distribution

Mixture of national distributions

Let $X_i, i = 1, ..., N$ be the income variable in the county i which is assumed to follow a GB2 distribution. The global CDF can be expressed as,

$$F(x) = \sum_{i=1}^{N} \lambda_i F_i(x) = \sum_{i=1}^{N} \lambda_i IB(v_i; p_i, q_i),$$

where λ_i stands for the population weights of the countries and $v_i = (x/b_i)^{a_i}/(1 + (x/b_i)^{a_i})$.

Global mean income

$$\mu = \sum_{i=1}^{N} \lambda_i \mu_i.$$

where μ_i is the per capita income of the country i

Data

1. Quintile data from UNU-WIDER World Income and Inequality Database (WIID), which is the longest and most comprehensive database of income distributions:

   https://www.wider.unu.edu/project/wiid-world-income-inequality-database

2. Top income shares from tax records come from the World Wealth and Income Database (WID)

3. Mean incomes per capita, per country-quantiles, are calculated based on GDP for the various country-years in 2005 US$ at PPP from the World Bank'

4. The number of individuals per country-quantile was calculated based on population data from the United Nations Population Division

What did we find?

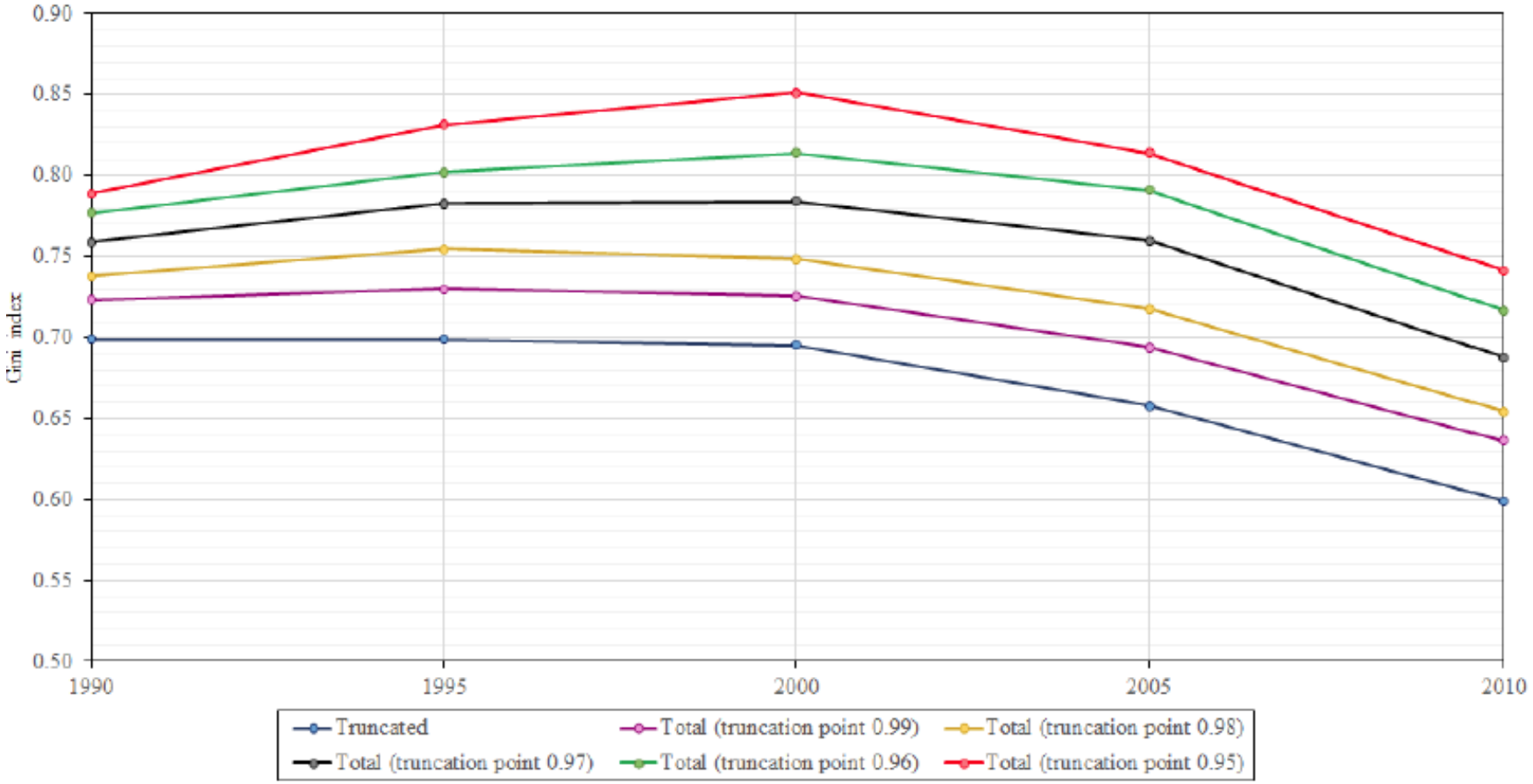
Global income inequality and estimated bias due to the omission of top income shares

		1990	1995	2000	2005	2010	1990-2000 Change (%)	2000-2010 Change (%)
MLD	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%
	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%
Theil	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%
	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%

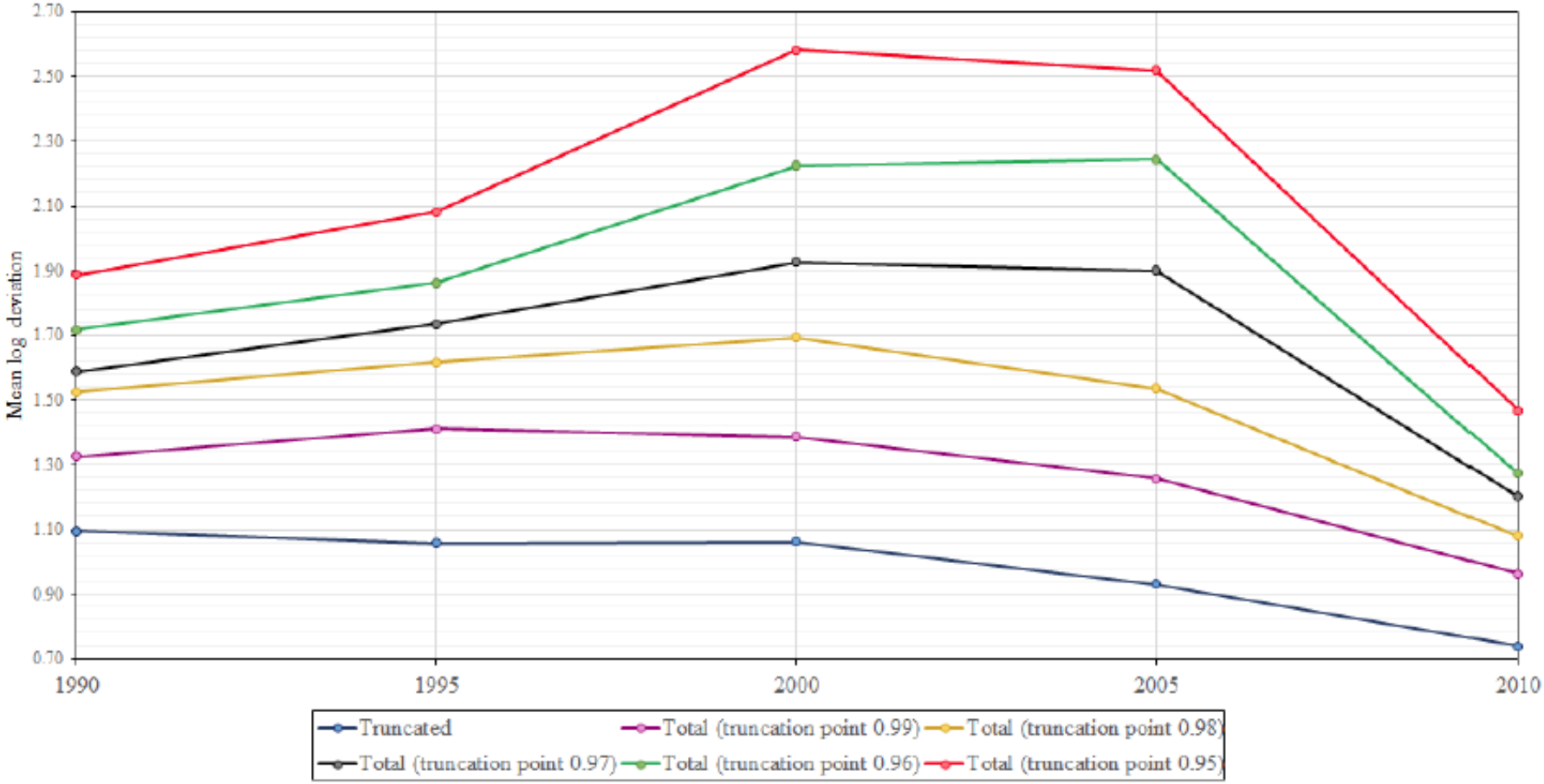
Decreasing trends of truncation points overtime

Undersampling the richest in HH surveys generated in 2010 a downward bias in global inequality that ranged between 17% and 38% using the MLD

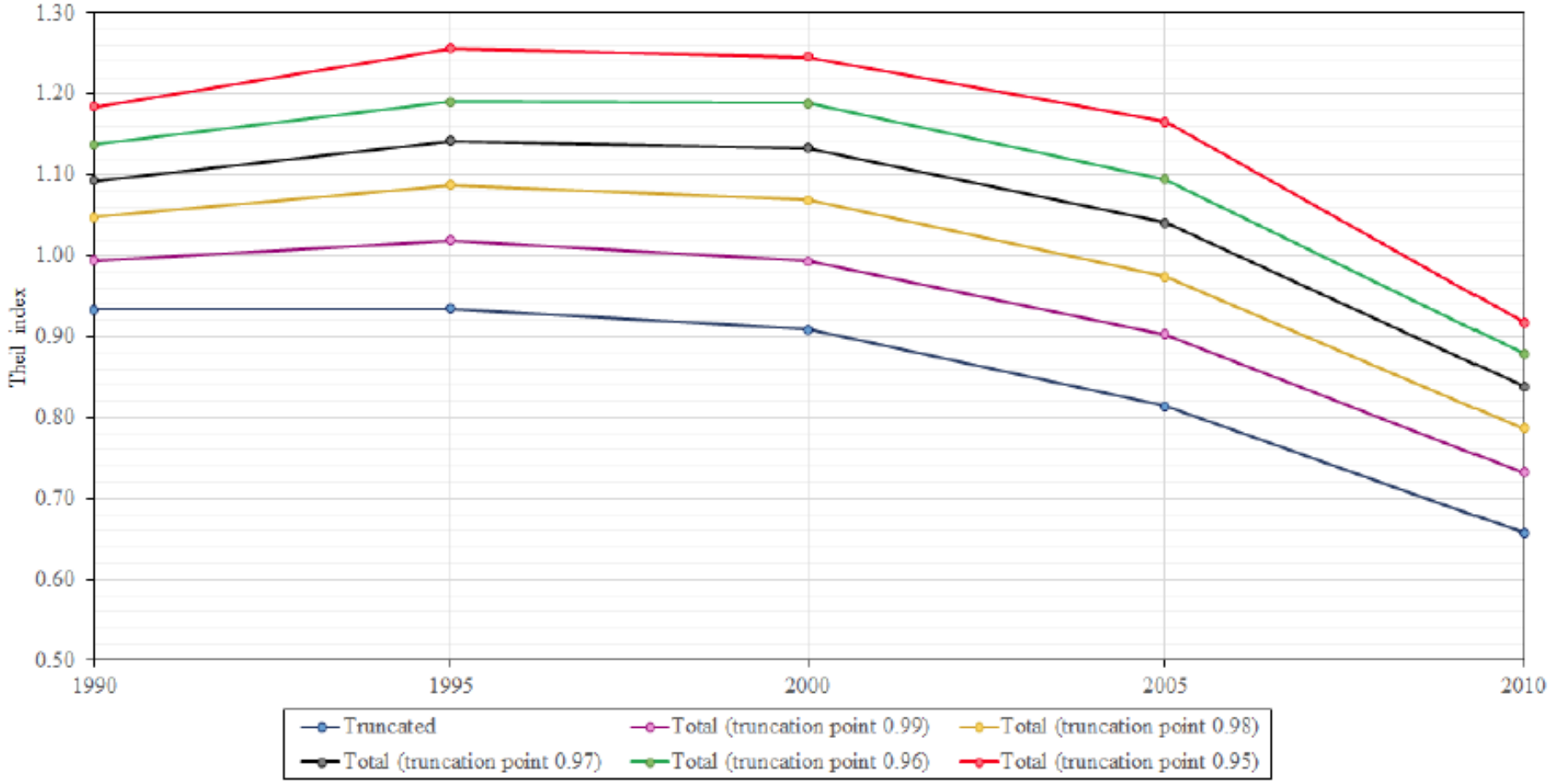
Global income inequality for different truncation points: Gini index



Global income inequality for different truncation points: MLD



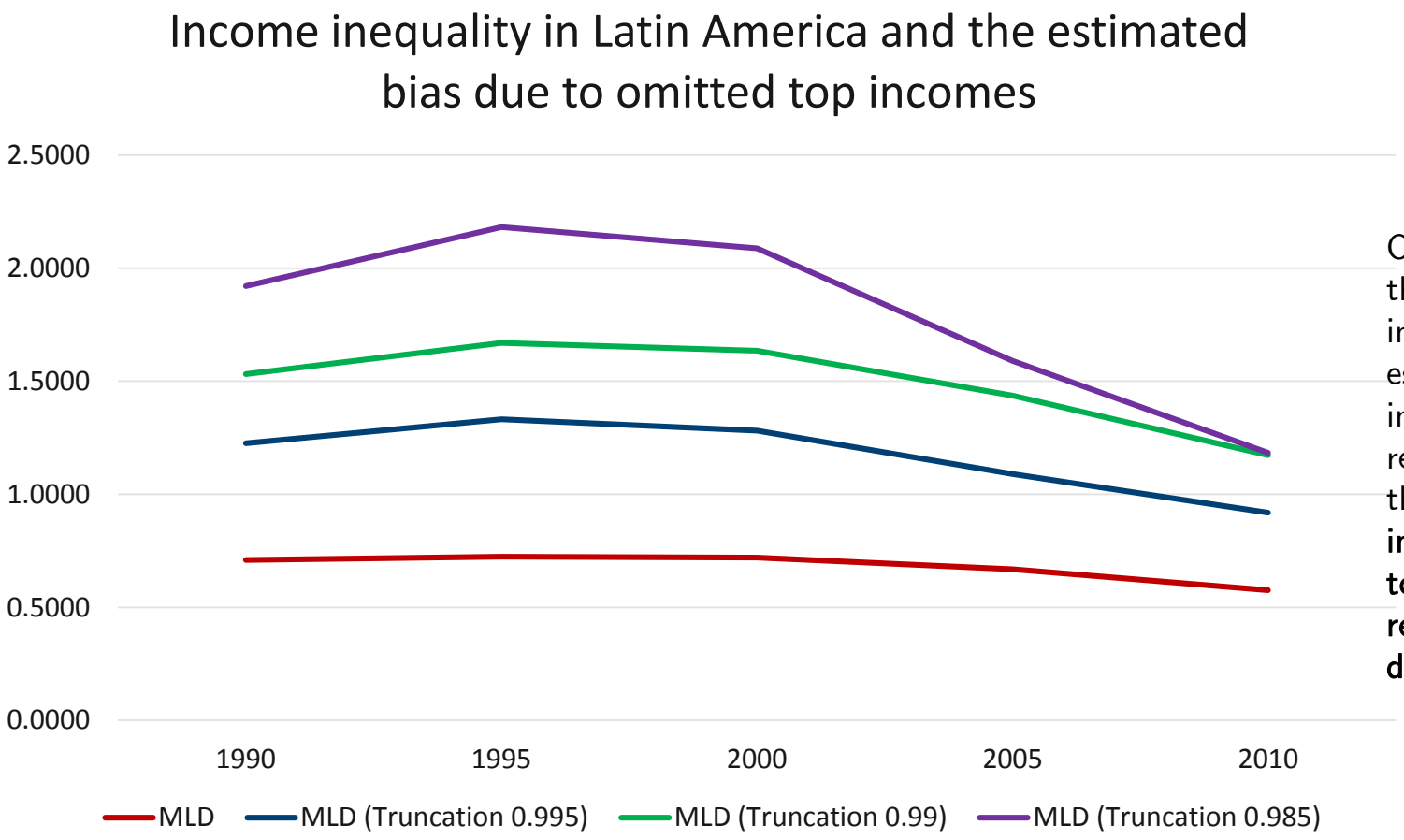
Global income inequality for different truncation points: Theil index



What happened at regional level?

- In contrast to what we observe at the global level, **within-country inequality is the predominant driver of regional inequality**
- **East Asia and the Pacific, the Middle East and North Africa** are the world regions that have seen changes not only in levels, but also in trends of inequality due to the imputation of top incomes
- **Europe 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
- **North America** displays an ascending trend in income inequality under conventional estimates for the period 2000-05, but this trend is reversed in the period 2005-10 after adjusting for top incomes bias
- This seems to confirm previous studies that show that the **richest top 1 per cent experienced the largest loss in income during and immediately after the financial crisis of 2008-09**, which in turn had a short-term `equalizing effect' in the income distribution (Alvaredo et al. 2013; Piketty and Saez 2013)

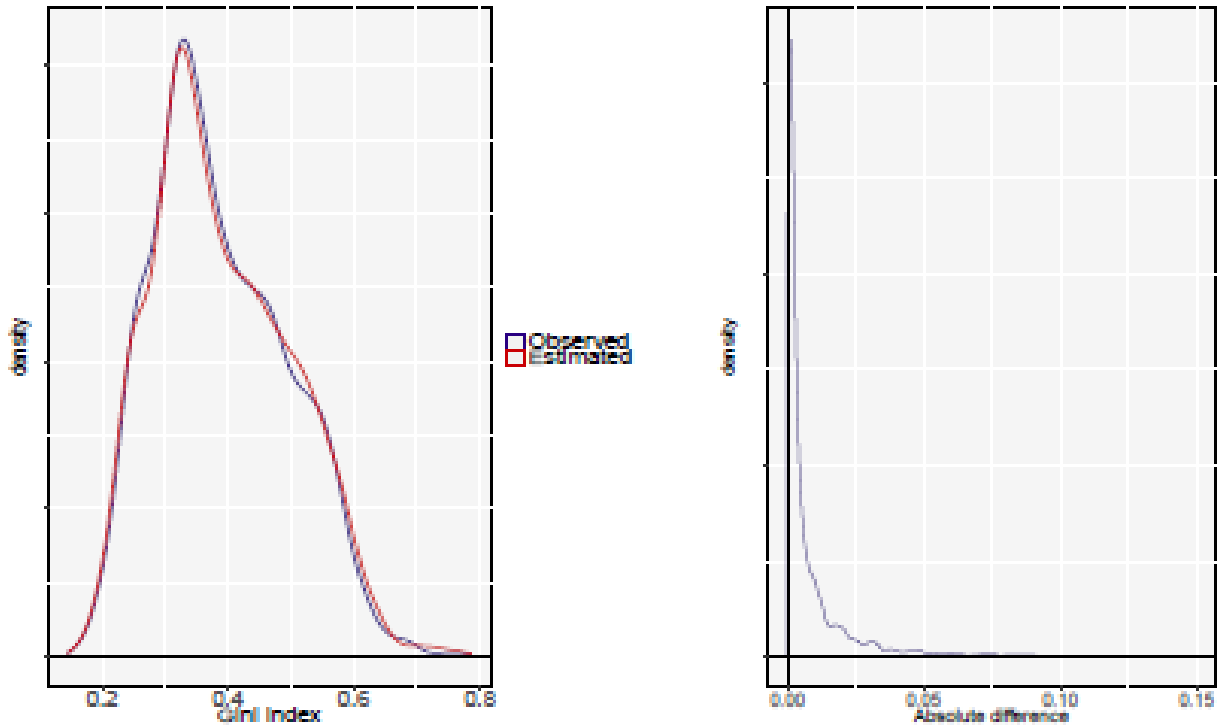
What has happened in Latin America?



Our results highlight the significance of top incomes in the estimation of income inequality and also reflect the titanic work that is needed to **improve accessibility to tax data for future research and policy design**

Goodness of fit

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



Concluding Remarks

- In this paper we estimate the levels and evolution of global and regional income inequality over the period 1990-2010, after accounting for the effect of omitted top incomes
- The proposed (GB2) methodology is a step forward towards improving our understanding of the impact of the richest on the evolution of global and regional inequality
- Our results indicate that survey-based analyses of global inequality suffer from **downward bias in the 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
- Our study highlights the importance of improving the coverage and accessibility to tax data for better policy design
- **Our results should still be treated as lower-bound estimates** given the fact that **under-reporting income bias** is likely to be present in tax records

www.wider.unu.edu

Helsinki, Finland