LSE

Francisco H. G. Ferreira

March 26th, 2024

Do we even know how much income inequality there is?

0 comments | 22 shares

Estimated reading time: 7 minutes

*There has been a proliferation of statistics that seek to capture the level of income inequality in countries and regions around the world. But what should we take from the staggering variation in the different measures that are used today, **Francisco H. G. Ferreira** asks? And what does this mean for the field of inequalities research?*

Interest in inequality has been rising steadily since the Global Financial Crisis of 2008/09. So much has been written about income inequality that you might expect scholars in the field to have a pretty good idea of how much of it there is and how it has been changing over time. Indeed,

there are public databases on the web that tell you precisely just how large inequality was in a particular country and year, for most countries in the world!

The problem is: they don't seem to agree on much. Figure 1 below, adapted from this recent paper, shows the income share of the richest 1% of the population for 53 countries in 2020. The horizontal axis shows the share reported in the World Income Inequality Database (WIID), maintained by the World Institute for Development Economics Research of the United Nations University. The vertical axis reports the share contained in the World Inequality Database (WID), maintained by World Inequality Lab at the Paris School of Economics.

Figure 1: Income share of top 1% by country

Source: wid.world/data and UNU-WIDER/WIIDcomp. All data for the year 2020
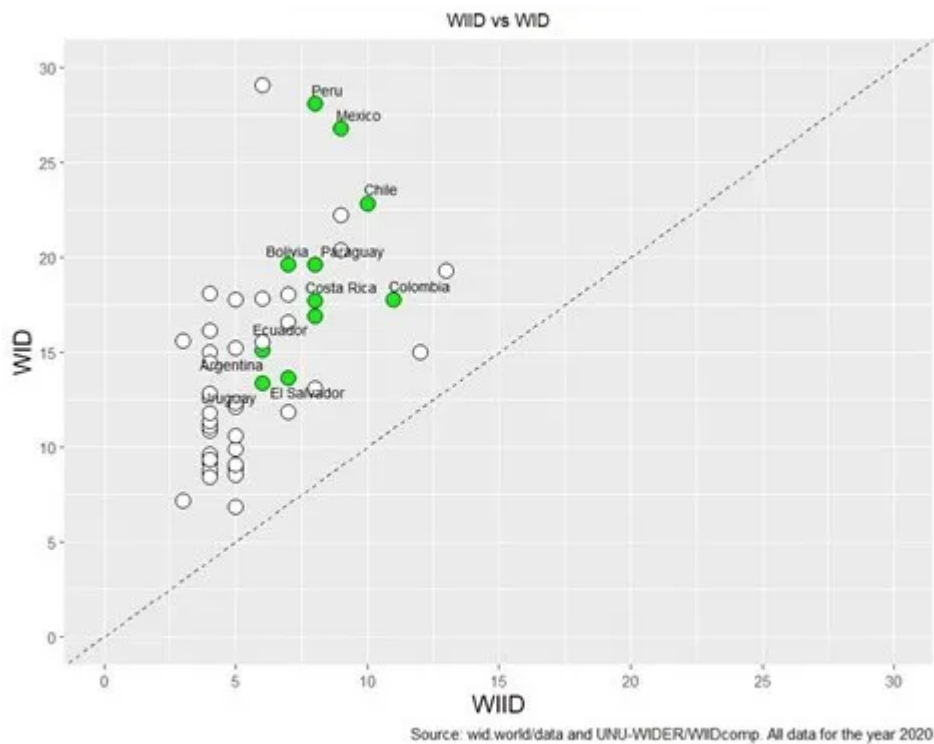
Both are well-reputed institutions, with considerable academic credibility. Yet, the WIID tells you that the richest 1% of Peruvians earn 8% of the country's total income, whereas the WID tells you the same group earns 28% of total income – in the same year! And although Peru is an extreme example, the pattern is widespread. The figure contains every country for which there were entries for 2020 in both databases

(as of July 2023). The average share reported in the WIID is 5.9%, and 14.3% in the WID. Clearly, they cannot both be right.

There is surprisingly little public discussion of these substantial discrepancies. Perhaps people (and journalists in particular…) have a low tolerance for uncertainty or nuance. A single 'truth' is typically preferred. In a recent paper with Facundo Alvaredo, François Bourguignon and Nora Lustig examining the history of inequality measurement in Latin America over the last seven decades, my co-authors and I compile all of the estimates of income inequality that we could find for the region, as measured by the Gini coefficient (a score of 0 represents perfect equality – everyone earning the same income – while a score of 1 represents perfect inequality, where one person earns all of the country's income).

We assembled over 5,600 estimates of the Gini coefficient, covering 34 countries in the region from 1948 to 2021, often with multiple estimates for the same country/year. Figure 2 illustrates our findings for Brazil, but it's a similar story for other countries with similar amounts of information. The disparity of estimates is staggering. In 1960, estimates of Brazil's Gini range from 0.43 to 0.70. In 2020, after decades of "progress", the range is not much narrower.

Figure 2: Gini coefficient estimates for Brazil from the 1950s to 2020
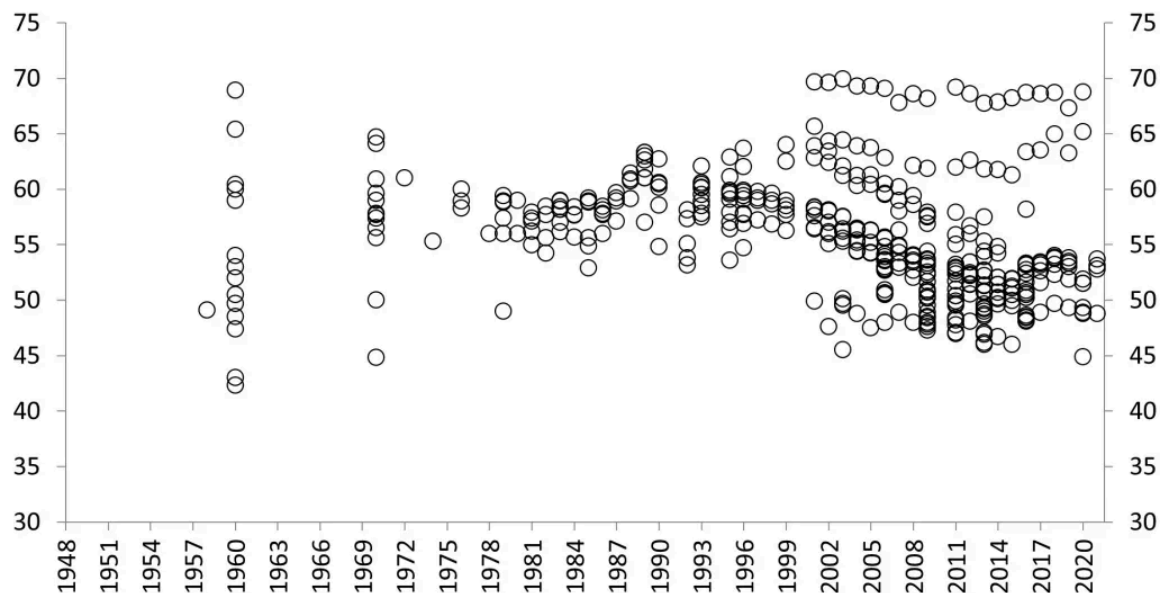
What explains the large differences in measured inequality?

In the paper, we discuss three categories of explanations.

(1) **Different data sources**. Researchers draw on various data sources – household surveys, censuses, tax records, social security data, National Accounts data – to calculate their estimates, sometimes combining information from multiple sources.

(2) **Different concepts (of income and populations)**. For instance, income could refer to individual earnings or household income; it could be taken net of taxes or not.

(3) **Differences in how the data is treated**. Even using the same concepts and data sources, differences may arise if some researchers correct for missing observations, for instance, while others do not, or if some trim the tails of the distribution, while others don't.

Using information about the various estimates included in our dataset, we assess the relative importance of these three sources of variation (we focus on a subset of ten countries for which we can conduct the most detailed analysis).

Looking at estimates for household per capita income across individuals based on household surveys alone, we find that the average annual range of Gini estimates is relatively small (1.7 percentage points). This suggests that Category 3 – different treatments of the same data – is not the main culprit. We find a wider range when we look at estimates still based only on household surveys but using various income concepts and population definitions: Category 2 seems to matter more than Category 3.

But it is Category 1 that accounts for the bulk of the disparity in inequality estimates in Latin America. For those ten countries, combining information from tax registries with household surveys raises the range to 10.4 percentage points. And when analysts try to distribute (across households) income that is "observed" in the National Accounts statistics but not in other sources, the average range rises to 16.9 points. That's a very wide range indeed!

Of course, if we were perfectly confident of each of these adjustments, the ranges would be indicative not of uncertainty, but of progress: better measurement would have raised our inequality estimates, largely by including top incomes and income from capital more generally, which were missed by the household surveys. This is, in fact, a commonly held view.

Unfortunately, there are reasons to be somewhat skeptical. It is certainly true that household surveys are woefully inadequate in accurately capturing the incomes of very rich people, and capital incomes more generally. On the other hand, in Latin America, as in other developing regions, large sections of the labour force are informal. Their incomes are much more likely to be captured by surveys than by tax or other administrative data.

Knowing this, a group at the World Inequality Lab has done some important work trying to combine information from the surveys and the

tax data. This was serious, careful research but, unfortunately, the proposed method is far from widely accepted, and others have found that the results it yields are highly sensitive to methodological choices which are, in turn, at least partly arbitrary. Things get even messier when attempts are made to distribute income from the National Accounts.

The ongoing work to refine measures of income inequality

Where does this leave us? On the one hand, the progress made in research on inequality over the last two decades or so – much of it led by Thomas Piketty, Emmanuel Saez, and their collaborators – has convincingly established that household surveys, on their own, tend to mismeasure top incomes and to underestimate income inequality. On the other, as we say in our paper, the adjustments which have been proposed to "correct" for that "involve a variety of assumptions and decisions which must be made on the basis of limited information, and to which results are quite sensitive."

We focused on Latin America, but recent debates about inequality in the US suggest that the problem may exist even in countries with richer data. In a paper forthcoming in the *Journal of Political Economy*, Gerald Auten and David Splinter take issue with various assumptions made earlier by Piketty and his co-authors, finding instead lower levels (and flatter trends) of inequality. Both papers will presumably have been carefully scrutinized and refereed. Yet they reach very different conclusions.

In our Latin America paper, we concluded that where levels of inequality are concerned, "we live in a world of some uncertainty, which we represent by means of a set of *inequality bands*". Similarly, discussing the US controversy, *The Economist* quotes Wojciech Kopczuk, of

Columbia University, as saying that "we are now getting (wide) bounds for where the truth might be".

It may not look like it, but this actually is progress! When most of us in this research field were looking only at household surveys, we thought we knew how much inequality there was, but we were typically underestimating it. Combining data from different sources, we have gained a more nuanced understanding of the income distribution. We may not yet be capable of exactness but, as the saying goes, it is better to be approximately right than precisely wrong…

All articles posted on this blog give the views of the author(s). They do not represent the position of LSE Inequalities, nor of the London School of Economics and Political Science.

Image credits: Photo by Clay Banks on Unsplash

About the author

Francisco H. G. Ferreira

Francisco H. G. Ferreira is the Amartya Sen Professor of Inequality Studies and Director of the International Inequalities Institute at the London School of Economics and Political Science. He is also affiliated with the Department of Social Policy at LSE.

Posted In: Income inequalities | Latin America inequalities