The Journal of Economic Inequality (2022) 20:223–243
https://doi.org/10.1007/s10888-021-09523-5

Declining inequality in Latin America? Robustness checks for Peru

Diego Winkelried[1] · **Bruno Escobar**[2]

Received: 5 January 2021 / Accepted: 13 December 2021 / Published online: 1 April 2022

Abstract
Household surveys underreport incomes from the upper tail of the distribution, affecting our assessment about inequality. This paper offers a tractable simulation method to deal with this situation in the absence of extra information (e.g., tax records). The core of the method is to draw pseudodata from a mixture between the income empirical distribution and a parametric model for the upper tail, that aggregate to a preestablished top income share. We illustrate the procedure using Peruvian surveys that, as in the rest of Latin America, have displayed a sustained decrease in the Gini index since the 2000s. In a number of experiments, we impose a larger top income share than the one observed in the data, closer to corrected estimates for less egalitarian neighbors (e.g., Colombia and Chile). We find that even though the point estimates of the Gini index are biased, the corrected indices still decrease in time.

Keywords Top income share · Income inequality · Latin America

1 Introduction

During the first two decades of the new millennium, the performance of Latin American countries in reducing poverty and income inequality was remarkable (see, *inter alia*, Lustig et al. 2013; Cord et al. 2015; Gasparini et al. 2011). According to our computations, the median Gini index of the region decreased 8 points in that period (from 53 to 45 points), significantly more than in any other region worldwide. These findings have fed an academic

We are indebted to Gustavo Yamada for insightful discussion. We would also like to thank Marco Terrones, Javier Torres, the editor Daniel Waldenström and two anonymous referees for very useful comments. The financial support of the Research Center of Universidad del Pacífico (Project PPA 2018) is gratefully acknowledged. We alone are responsible for the views expressed in this paper and for any remaining errors.

✉ Diego Winkelried
winkelried_dm@up.edu.pe

Bruno Escobar
b.escobarizquierdo@up.edu.pe

[1] School of Economics and Finance, Universidad del Pacífico, Av. Salaverry 2020, Lima 11, Peru

[2] School of Economics and Finance, Universidad del Pacífico, Lima, Peru

debate on shared prosperity and the pro-poor nature of the high economic growth rates sustained in the region (see, *inter alia*, Gasparini et al. 2011; Cord et al. 2015; Amarante 2016).

As it is often the case in the study of income dynamics in developing countries, this evidence is mostly based on household living conditions surveys (see, *inter alia*, Bourguignon 2015; Stampini et al. 2016). The literature recognizes that such surveys tend to underestimate top income shares and inequality indices due to the systematic undercoverage of richer households and underreporting of their total income (Atkinson 2007; Deaton 2005; Lustig 2019).[1] A common practice to correct these issues – following the seminal works of, *inter alia*, Piketty (2003), Piketty and Saez (2003) and Atkinson (2005) – is to complement survey data with administrative records of income taxes that measure the upper tail of the income distribution better (see also Atkinson et al. 2011). In Latin America, corrections of survey statistics are available for Argentina (Alvaredo 2010, 2011), Brazil (Assouad et al. 2018; Morgan 2018), Chile (Flores et al. 2020), Colombia (Alvaredo and Londoño Vélez 2013; Díaz-Bazán 2015) and Uruguay (Higgins et al. 2018; Burdín et al. 2020). See De Rosa et al. (2020) for more recent estimates.

However, the corrections made to inequality indices are not without criticism and may still be subject to some degree of uncertainty (see Saez and Zucman 2020). The corrected point estimates can themselves be upward biased due to multiple and necessary specification choices made during imputation.[2] Hence, some authors even call for a shift in paradigm to focus on ranges rather than point estimates, recognizing the imperfections embedded in all methods to accurately estimate inequality (Burkhauser et al. 2015; Jenkins 2017; Auten and Splinter 2019; Lustig 2019; Aaberge et al. 2020).

Along those lines, this paper offers an accessible simulation method to assess the mismeasurement of the upper tail of the income distribution and to subsequently adjust the inferences made about income inequality. Importantly, our methods works on survey data alone, and does not use additional administrative or tax data. Borrowing ideas from the semiparametric bootstrap scheme advanced in Cowell and Flachaire (2007) and Davidson and Flachaire (2007), the purpose is to draw random numbers from a corrected income distribution, which is a mixture between the empirical distribution of the household survey and a suitably parameterized Pareto model for the top incomes (see, for instance, Reed 2003; Atkinson 2007). Then, these draws constitute pseudodata that on average do not underestimate top income shares, which can help to track the evolution of corrected, or 'true', inequality measures and its uncertainty.[3]

The procedure depends on a single parameter that connects both parts of the mixture: the unobserved 'top income share', i.e., the true share in total income of the households in the top income group. This share, which is a main object of study in the literature (see Ruiz and Woloszko 2016; Anand and Segal 2017, for reviews), is calibrated based on educated

[1] As pointed out in Higgins et al. (2018) and Lustig (2019), undercoverage stems from income item and unit non-response, while underreporting of covered units is a well-documented fact among rich households.

[2] For an illustrative example, Saez and Zucman (2020) revise their previous estimations for inequality in the US, for instance in Piketty et al. (2018), based upon legitimate caveats pointed about their original methodology. These include concerns about their assumptions used to impute S-corporation profits as capital income (Smith et al. 2019), the estimates of private business wealth (Smith et al. 2021), and the allocation of untaxed pension income (Auten and Splinter 2019). After addressing these issues, their estimates still show a considerable increase in inequality, but not as steep as originally shown in Piketty et al. (2018).

[3] In terms of the classifications in Lustig (2019), ours is a 'replacing method' where the top income statistics are produced by artificial data whose characteristics, such as coming from a heavy-tailed distribution, can mimic the adjustments made by methods that use external information such as tax data.

guesses that treat survey computations as lower bounds, and corrected estimates from the studies of countries believed to be less egalitarian as upper bounds. Thus, given the survey data, the simulations exploit a one-to-one map from the top income share to inequality measures – mainly, the Gini index – thereby unveiling transparently how different beliefs about the misrepresented top income tail affect our assessment of income inequality, and vice versa.

We illustrate the method with Peruvian national household survey data from 2004 to 2018. Comparative studies, such as Ferreira et al. (2013) or Cord et al. (2015), regard the Peruvian experience as successful, and we document that it is representative of the regional tendency. Moreover, the survey top 1% income share is less than 10 percent which is less than half the corrected share reported for neighbor countries such as Colombia or Chile, that is about 20 and 25 percent, respectively. By using 20 percent as the top income share of the unobserved true income distribution, we could obtain the joint distributions of adjusted Gini indices to assess both the magnitude of the bias in the surveys and the probability of reductions in the true indices. We conclude that even though the corrected Gini indices are larger than the uncorrected ones, they are also significantly decreasing throughout the sample period.

The rest of the paper is organized as follows. Section 2 gives a short comparative review of the Latin American experience in reducing inequality. Then, it presents a review of the top income literature. Section 3 presents the methodological framework, and provides a step-by-step description of the algorithm proposed to draw incomes from a distribution with a preestablished top income share. Section 4 contains our empirical exploration. Finally, Section 5 offers closing remarks and some avenues for future research.

2 Background

We present a brief account of the recent decline of inequality in Latin America. The evidence supporting this phenomenon does not correct for biases in the top income shares, which is a topic we also discuss below.

2.1 Declining inequality in Latin America

Widely regarded as one of the most unequal regions in the world during the 20th century, Latin America has shown remarkable progress in reducing income inequality in the past 20 years. These dynamics and their determinants are extensively documented elsewhere (see, *inter alia*, Gasparini 2004; Gasparini et al. 2011; Lustig et al. 2013; Cord et al. 2015; Amarante 2016; Stampini et al. 2016; Székely and Mendoza 2016), so we just provide a concise description below. To this end, we use the compilation of Gini indices in the World Development Indicators (`WDI`) database, that in turn use comparable country-specific household surveys as the primary source of information.[4]

Figure 1a shows the Gini indices by 2018 (averages between 2016 and 2018) against the indices by 2000 (averages between 2000 and 2002) for 67 countries around the world: 15

[4]The code for the Gini index in the `WDI` is `SI.POV.GINI`. An alternative resource for the cross-country analysis of income inequality are survey-based measures compiled by the United Nations Economic Commission for the Latin American and the Caribbean (ECLAC). Bourguignon (2015) presents a comparison between these sources, whereas Gasparini (2004) and Gasparini et al. (2011) show that although it is possible to find country-specific discrepancies, the region-wide dynamics are close in both databases.

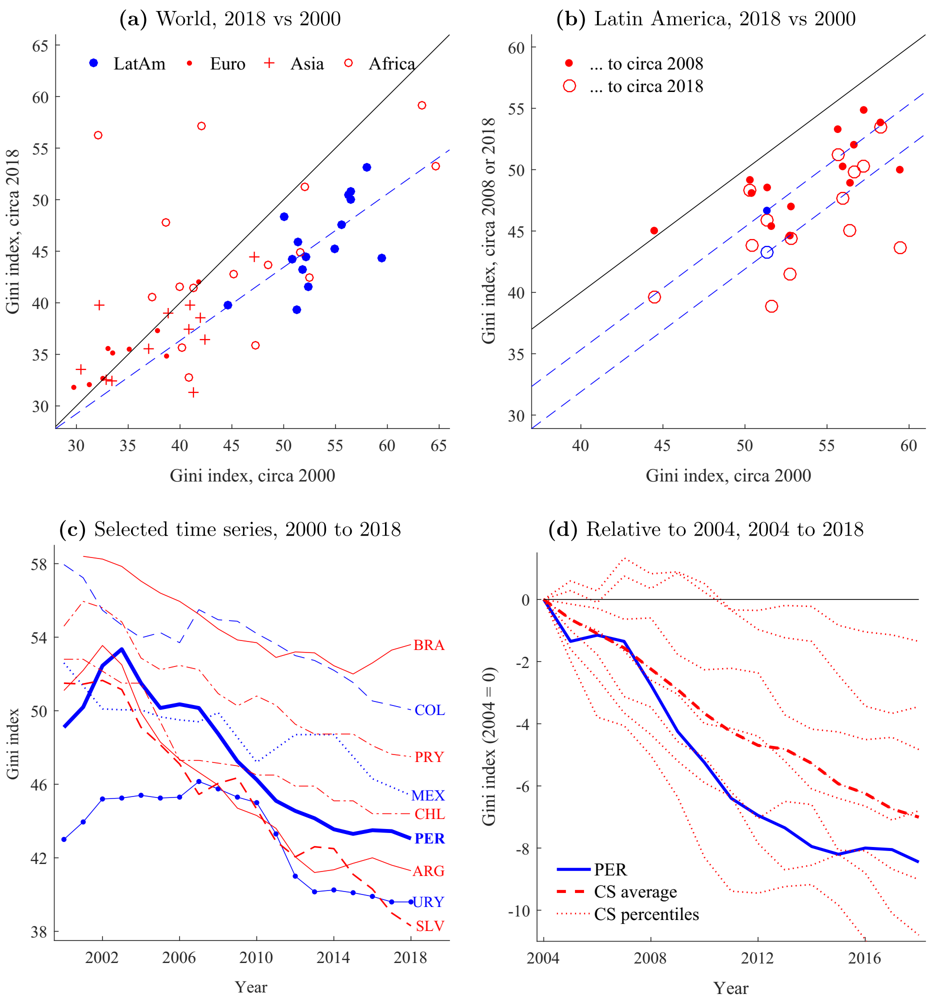

Fig. 1 Gini indices in the world and in Latin America, 2000 to 2018 **Source:** World Bank, World Development Indicators. Own elaboration. **Notes:** Panel (a): Gini indices in 2016/2018 vs 2000/2002 for 67 countries worldwide. The continuous line is a 45-degree line; the dashed line in the regression among Latin American countries. Panel (b): Gini indices in 2016/2018 and 2008/2010 vs 2000/2002 for 14 Latin American countries. The dashed lines pass through the Peruvian coordinates to mark the vertical distance from the equality (continuous) line. Panel (c): 2-year center moving averages of the Gini indices; missing values replaced by linear interpolations. Panel (d): Gini indices minus their values in 2004; the dashed lines are cross-sectional percentiles (5, 10, 25, 50, 75, 90 and 95)

Latin American, 14 European, 12 Asian and 16 African. By the beginning of the period, the median Gini index for the European, Asian and African countries is, respectively, 32, 40 and 44 points.[5] For Latin America it is 53 points, being initially the most unequal region in our sample. The continuous line is an equality (45-degree) line, so the points below correspond

[5]The high volatility of the African data is due to the diversity of inequality patterns thoroughly studied in Chancel et al. (2019).

to countries where the Gini index decreased between 2000 and 2018. This is the case for all Latin American countries, with a decrease in the median of 8 points that is significantly higher than the 3 and 1 points in Asia and Africa, respectively, and far from the 0.5 point increase in Europe. By the end of the period, the median index in Latin America is 45 points, similar to that of Africa.

Figure 1b focuses on Latin America, and shows the comparison between the initial (average from 2000 to 2002) and final indices (average from 2016 to 2018, hollow circles), and an intermediate situation (average from 2008 to 2010, filled dots). The reduction in the Gini index is not only pervasive but also sustained and evenly paced: the median index is reduced from 53 points in 2000 to 49 points in 2008 (a 4 point decrease), to 45 points (a further 4 point decrease).

The dashed lines mark the vertical distances with respect to the coordinates of Peru, our case study below. Points above [below] these lines correspond to countries with a smaller [higher] decrease in the Gini index than Peru. The Peruvian experience is considered quite successful in poverty reduction and the decrease of inequality (see, *inter alia*, Ferreira et al. 2013; Genoni and Salazar 2015; Paz and Urrutia 2015; Herrera 2017; Winkelried and Torres 2019). From a regional perspective, however, the Peruvian dynamics are representative of the regional central tendency, with some countries performing worse and some others performing better.

Alternatively, Fig. 1c shows the sustained decrease in all Gini indices, as the evolution of time series. Peru is among the best performing countries, while still not being an outlier. Finally, Fig. 1d shows the evolution of the Gini index relative to 2004, the initial period in our analysis below, for Peru and cross-sectional percentiles of the remaining countries. The cumulative reduction in the Peruvian Gini index is about 8 points, slightly higher than the median (and average) reduction.

2.2 Top incomes and household surveys

It has been well established that top income share estimates based solely on living standards surveys tend to be underestimated (see, *inter alia* Burkhauser et al. 2012; Anand and Segal 2017). Lustig (2019) argues that the difficulty arises due to a number of non-excluding factors: namely, the sampling frame could exclude high-income neighborhoods by design; the pollster may not be granted access to exclusive areas; high-income respondents could be more prone to reject being surveyed altogether or to refuse answering questions related to income declaration. These issues relate to unit and item non-response and lead to undercoverage of the upper incomes. However, even when respondents do answer, households in the top income groups may consistently underreport their income.[6] The combination of undercoverage and underreporting factors cause sparseness (i.e., the lack of density mass) of income in the upper tail and, in severe cases, right truncation (Jenkins 2017).[7]

[6] More generally, households and individuals may misreport their income either downwards or upwards. However, as stressed by Higgins et al. (2018) and Lustig (2019), the comparison of survey data with tax data made it quite apparent that the predominant bias comes from people at the top of the income distribution systematically underreporting their income, especially income from capital.

[7] Another possibility, albeit not as common as truncation, is right-censoring in income data. Burkhauser et al. (2012) provides a comprehensive discussion for the US. See also Lustig (2019) for further reference.

As a result, a growing body of research has developed several methods to correct survey-based statistics (see, *inter alia* Atkinson et al. 2011; Ruiz and Woloszko 2016). A first branch complements the data used in the assessment of inequality with additional sources. The pioneering works of Piketty (2003) for France, Piketty and Saez (2003) for the US and Atkinson (2005) for the UK, use income data derived from public tax records instead, as this source of information admittedly capture top incomes much better. Nonetheless, as argued in Atkinson (2007) and Atkinson et al. (2011), tax data are not as representative as survey data for the lower end of the income distribution. Thus, Atkinson (2007) proposes to compute inequality indices by combining estimates for the upper tail from the tax data with estimates for the remainder of the population from survey data. This methodology has been adopted to study inequality in the US (Alvaredo 2011; Atkinson et al. 2011) and mainly developed countries, from Europe (Burkhauser et al. 2016; Jenkins 2017; Aaberge et al. 2020) and others compiled in Atkinson and Piketty (2010).

A second approach to correct inequality estimates relies on readily available national accounts. In its simplest form, the idea is to impute the differential between the survey mean income per unit of observation and per capita income from the national accounts to the share of income of the unobserved upper tail.[8] With this approach, Lakner and Milanovic (2016) and Anand and Segal (2017) revise the Gini indices on a global scale. However, two drawbacks are that the different definitions of income in surveys and national accounts render them incomparable (see Yamada et al. 2012; Jenkins 2017), whereas unavoidable measurement errors and revisions in national accounts result in sizable and artificial changes in inequality estimates (see Deaton 2005; Bourguignon 2015).

A compromise of the aforementioned methods, the so-called 'Distributional National Accounts' (DINA) series popularized in Piketty et al. (2018) has grown quite popular in the field of income and wealth inequality estimations. The guidelines by Alvaredo et al. (2020) offer a comprehensive explanation of the method; we summarize some key points here. A DINA income series combines survey, tax and other income datasets (e.g. undistributed corporation profits or indirect taxes) to estimate an income distribution that distributes the entirety of net national income – as recorded in the National Accounts – across all resident adult population. After the harmonization of income datasets from different countries (with PPP exchange rates) and through time (with price indices), a DINA panel provides a complete description of the distribution from the bottom to the very top of a geographic region, such as Latin America.

DINA series currently form the main source of information for the comprehensive income distribution statistics that are publicly available in the World Inequality Database (`WID.World`). Throughout the paper, we use data of corrected top incomes shares from `WID.World`, but the database offers a wealth of additional information, including income shares from multiple subgroups in a country or regions (e.g. bottom 50%, middle 40%, top 10%, top 0.01%) for several income definitions, wealth distribution statistics, and the key macroeconomic aggregates employed to compute DINA estimates. The country selection is also quite generous: besides Europe and the US (Blanchet et al. 2019), the list includes countries in regions such as Africa (Chancel et al. 2019) and the Middle East (Alvaredo et al. 2019), and individual countries such as China (Piketty et al. 2019), India (Chancel and Piketty 2019), Russia (Novokmet et al. 2018) and South Korea (Kim 2018). For Latin

[8] The unit of observation most commonly used in the literature is the individual. When the income record is at the household level, Alvaredo et al. (2020) suggest splitting it evenly among household members to obtain an approximate of individual-level income.

American, the selection includes Argentina (Alvaredo 2010, 2011), Brazil (Morgan 2018; Assouad et al. 2018), Chile (Flores et al. 2020), Colombia (Alvaredo and Londoño Vélez 2013; Díaz-Bazán 2015) and Uruguay (Higgins et al. 2018; Burdín et al. 2020).

Figure 2 reports `WID.World` top 1% income shares for a diverse selection of countries, for around the early 2000s and around the late 2010s. The top 1% shares can be about 10 percent in relatively more egalitarian economies, about 20 percent in highly unequal countries and between 25 and 30 percent in extremely unequal countries (see Assouad et al. 2018). In general, these shares remain stable, but there are some noticeable increases: namely, Central African Republic in Africa, Qatar in the Middle East, Chile in Latin America, India for large emerging markets, and the US and South Korea among advanced economies.

Figure 2 reports `WID.World` top 1% income shares for a diverse selection of countries, for around the early 2000s and around the late 2010s. The database was recently updated in late 2020 to include a comprehensive time series for most countries worldwide, and every country in our sample selection. The top 1% shares can be about 10 percent in relatively more egalitarian economies, about 20 percent in highly unequal countries and between 25 and 30 percent in extremely unequal countries (see Assouad et al. 2018). In general, these shares remain stable, but there are some noticeable increases: namely, Central African Republic in Africa, Chile and Brazil in Latin America, India for large emerging markets, and the US and South Korea among advanced economies.

Notwithstanding its popularity, tax-based and DINA methods are not free from critiques. As illustrated in the back and forth discussion and revision of inequality estimates between Auten and Splinter (2019), Smith et al. (2019, 2021) and Piketty et al. (2018) and Saez and Zucman (2020), these methods can be quite sensitive to the assumptions made to combine the information from three different sources. Consequently, authors such as Lustig (2019) and Aaberge et al. (2020) suggest to produce lower and upper bounds of inequality measures, rather than point estimates.

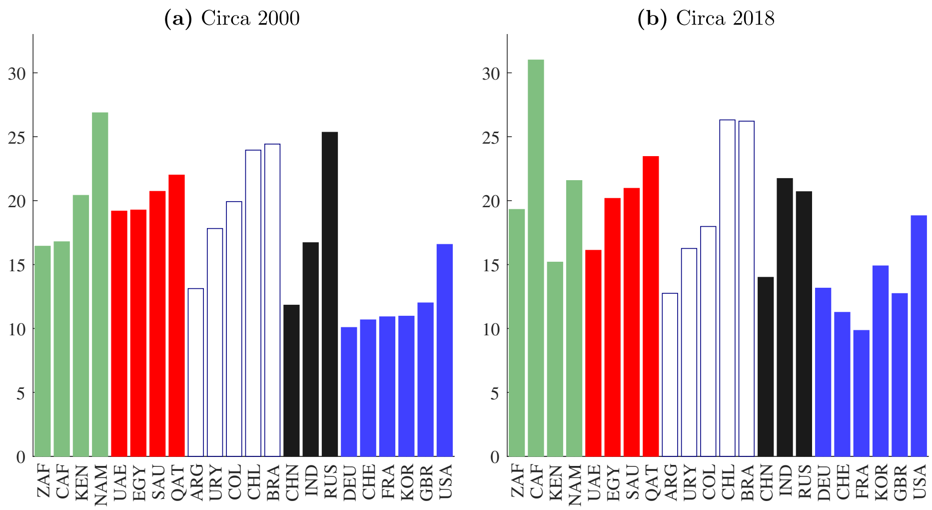

Fig. 2 Top 1% income shares for selected countries, 2000 to 2018 **Source:** `WID.World`, November 2020 update. Own elaboration. **Notes:** Panel (a) presents the average top 1% income shares between 2000 and 2003; panel (b), between 2015 and 2018. In the `WID.World` database, income is before taxes. The shares are sorted within each group (Africa, the Middle East, Latin America, large emerging economies and OECD countries), according to the initial values

Another branch of the literature does not rely on additional data. Rather, it aims to adjust the statistical and inferential procedures to address the intricacies of the upper tail, such as sparseness and undersampling. The methods in Cowell and Flachaire (2007) and Davidson and Flachaire (2007) correct for the sparseness in the high end of the sample on inequality indices by replacing the actual top income data with synthetic observations from appropriately fitted parametric distributions. Ruiz and Woloszko (2016) propose corrections in the same spirit by fitting a Pareto distribution to the top income data. A different but related approach is advanced in Alfons et al. (2013), who propose weighted estimators to compensate for the under-reporting of top incomes (see also Charpentier and Flachaire 2019). Eckerstorfer et al. (2016), who focus on wealth data, also correct the data after interpreting the downward bias as a natural consequence of non-observability in small samples drawn from a skewed distribution.[9]

Our simulation method draws heavily from these methodological advances, but with the important difference that it does not aim to exploit the survey data to improve our predictions of the upper tail or to correct it. Despite the refinements delivered by these methods, studies such as Burkhauser et al. (2012) and Higgins et al. (2018) argue that, depending on the degree of underestimation, corrected estimates based exclusively on household survey data may remain downward biased. Instead, we propose a framework for comprehensive sensitivity analyses on how inequality measures – mainly, the Gini index – respond to educated guesses or calibrations of the top income share. The outputs are bounds and probability statements about the corrections and changes of the survey-based Gini indices.

An important precedent to our work is Blanchet et al. (2018). These authors also argue that there is no convention on a standardized procedure to merge income and tax data, and develop a correction methodology that, unlike ours, use tax data but, like ours, use simulations in artificial populations with controlled misreporting and non-response rates. Compared to theirs, our approach is simpler. Also, we assume that non-response and mismeasurement of income happens exclusively in the top income group, and that no underestimation of consequence occurs in rest of the population.

3 Methodological discussion

This section presents the theoretical framework behind our simulations. The purpose is to develop an algorithm to generate samples from a distribution with the characteristics of the survey data, but that also incorporates adjustments in the upper tail such that the resulting draws aggregate to a preestablished top income share. Embedded in the algorithm is the assumption that the underestimation of income is more material at the top income group.

3.1 Pareto model for top incomes

Following a long tradition in the study of income distributions, we use a Pareto distribution to model the upper tail of the income distribution, i.e. the distribution of all incomes greater than the threshold Y_P. Despite its simplicity, this distribution is appealing from a theoretical viewpoint (Reed 2003) and has received enormous empirical support (see Clementi and

[9] Vermeulen (2018) develops an innovative approach that involves both reweighting and the use of information from 'The World's Billionaires' ranking of Forbes magazine to adjust for the far upper tail of the wealth distribution. Forbes' ranking does not provide income data, though.

Gallegati 2005; Cowell and Flachaire 2007; Charpentier and Flachaire 2019) both as an exact model for the upper tail, or as a useful approximation.[10]

A continuous random variable $Y \geq Y_P$ follows a Pareto distribution with "tail index" $\alpha > 1$, if its survival function is given by:

$$1 - F(Y) = \left(\frac{Y_P}{Y}\right)^{\alpha}. \tag{1}$$

As explained in Ibragimov and Ibragimov (2018), the tail index α measures the rate of decay or the heaviness of the tails which, in turn, governs the likelihood of observing large observations and fluctuations in Y. A smaller value of α corresponds to a heavier tail and, thus, to a higher probability of extreme values. Thus, not only dispersion measures, such as the Gini index or the variance (finite for $\alpha > 2$), but also the mean are decreasing functions of α:

$$\text{Gini}(Y \mid Y \geq Y_P) = \frac{1}{2\alpha - 1}, \quad \text{E}(Y \mid Y \geq Y_P) = \left(\frac{\alpha}{\alpha - 1}\right) Y_P \tag{2}$$

$$\text{and} \quad \text{var}(Y \mid Y \geq Y_P) = \frac{\alpha}{\alpha - 2}\left(\frac{Y_P}{\alpha - 1}\right)^2. \tag{3}$$

From Eq. 1, random draws from the Pareto distribution are easy to generate with an inverse transformation sampling method. Namely, $Y = Y_P / w^{1/\alpha}$ where w is uniformly distributed on $(0, 1)$.

3.2 Gini index and the top income share

The Gini index is additively decomposable when income does not overlap among members of different population groups. This is the case, for example, when the population is classified into mutually exclusive income classes (see, *inter alia*, Atkinson 2007; Jenkins 2017). For the two-class case, Alvaredo (2011) shows that the Gini index in the whole population, G, can be written as:

$$G = P\,S\,G^{**} + (1 - P)(1 - S)G^{*} + (S - P). \tag{4}$$

where P is the proportion of the population in the top income group; S is the top income share; G^{**} is the Gini index among households in the top income group; and G^{*} is the Gini index among households in the rest of the population or non-top income group. The sum of the first two terms in Eq. 4 is the within-groups component, whereas the last term, that takes the very simple form $S - P$, is the between-groups component. When $P \to 0$ and S remains finite, G can be readily approximated by:

$$G = (1 - S)G^{*} + S = G^{*} + (1 - G^{*})S, \tag{5}$$

The share S plays a key role in our analysis. Consider that, as usual, $G^{*} \in (0, 1)$. From Eq. 5), it follows that G is strictly increasing in S, ranging from the lower bound of $G = G^{*}$ (for $S = 0$) to the maximum admissible value of $G = 1$ (for $S = 1$). Put differently, the added inequality due to the presence of the top income group, $G - G^{*} = (1 - G^{*})S$, is increasing in S.

Regarding the correction of sample statistics, recall that G^{*} can be precisely computed from survey data, which also provides an estimate of the top income share $\underline{S}$ that

[10] A classic result, the second theorem of extreme value theory due to Balkema and De Haan (1974), states that every fat-tailed distribution will tend to behave as a Pareto distribution as Y_P increases.

is biased downwards and, correspondingly, a smaller total population Gini index of $\underline{G} = G^* + (1 - G^*)\underline{S}$. It follows that $G - \underline{G} = (1 - G^*)(S - \underline{S})$, so when the survey information underestimates S, then it also underestimates G: $G > \underline{G}$ if and only if $S > \underline{S}$. An implication is that any correction of the Gini index for the effects of the top income group should necessarily correct the top income share.[11]

Furthermore, if we consider the Pareto model for the top income group, we can build an invertible map from α to S. Denote the mean income in the top income group as $\mu^{**} = \alpha Y_P/(\alpha - 1)$, and let μ^* be the mean income in the rest of the population, which can be directly computed from survey data. The top income share can be written as $S = P\mu^{**}/(P\mu^{**} + (1 - P)\mu^*)$. In terms of the parameters of the Pareto model:[12]

$$S = \left[1 + \rho\left(1 - \frac{1}{\alpha}\right)\right]^{-1} \quad \text{where} \quad \rho = \left(\frac{1 - P}{P}\right)\frac{\mu^*}{Y_P}. \tag{6}$$

For $\alpha > 1$, this is a strictly decreasing function of α. When $\alpha \to 1$ from the right, then $S \to 1$. On the other hand, as $\alpha \to \infty$, S converges to the finite value $(1 + \rho)^{-1}$. It follows that the Gini index of the whole population G is also a decreasing function of α, ranging from $G = 1$ when $\alpha \to 1$ from the right, to G close to G^* when $\alpha \to \infty$. The inverse function of Eq. 6 is:

$$\alpha = \left[1 - \frac{1}{\rho}\left(\frac{1}{S} - 1\right)\right]^{-1}. \tag{7}$$

3.3 Sampling from a mixture with a Pareto upper tail

Consider a vector of income from survey data $\mathbf{Y}$, sorted increasingly. We develop a simple algorithm to draw observations from the mixture of a Pareto model for the top income group ($Y \geq Y_P$), and the empirical distribution function of $\mathbf{Y}$ for the rest of the population ($0 \leq Y < Y_P$). Moreover, since ρ is obtained directly from $\mathbf{Y}$, the tail index can be computed through Eq. 7 for a choice of S. Therefore, the algorithm produces a vector of pseudo data $\mathbf{y}$ with an expected share S, which is taken as one of the simulation parameters.[13]

Any consistent estimator of the Gini index that use the data $\mathbf{y}$ will converge in probability to the corrected G, as decomposed in Eq. 4). Moreover, the simulator can be used to compute bounds or probabilities of events related to inequality over time. In particular, consider the data of two surveys at periods 1 and 2, $\mathbf{Y}_t$ with N_t observations for $t = \{1, 2\}$ and true Gini index G_t. The b-th replication of the simulator will produce the estimates $G_t^{(b)} = G_t^{(b)}(\mathbf{y}_t)$ for $t = \{1, 2\}$. Then, the probability that $c_1G_1 + c_2G_2 > c_3$, for fixed c_1, c_2 and c_3 can be

[11] That G is strictly increasing in S also holds when P is not infinitesimal. The expressions involved in the general case are tedious and do not provide any additional insight to Eq. 5.

[12] By construction, Y_P increases as P decreases. Thus, the combination of an infinitesimal P but with a finite S can be achieved by keeping PY_P fixed as $P \to 0$. In this case, Eq. 6 holds with $\rho = \mu^*/(PY_P)$.

[13] An alternative route is to adopt a fully parametric function from the Pareto-lognormal family to model the complete income distribution, and to draw $\mathbf{y}$ from it. See Reed and Jorgensen (2004) and Bee (2015) for recent developments. The advantage is that many equality measures are available in closed-form and that the parametric nature of the approach can render efficiency gains (see Hajargasht and Griffiths 2013). A drawback is that the estimation of Pareto-lognormal parameters can be challenging. We believe that exploring further this parametric models is an interesting topic for future research.

well approximated by a crude frequency counting:

$$\frac{1}{B}\sum_{b=1}^{B} \mathrm{I}(c_1 G_1^{(b)} + c_2 G_2^{(b)} > c_3) \xrightarrow{p} \Pr(c_1 G_1 + c_2 G_2 > c_3) \quad \text{as} \quad B \to \infty\,, \tag{8}$$

where $\mathrm{I}(x)$ is the step function such that $\mathrm{I}(x) = 1$ if x is true, and $\mathrm{I}(x) = 0$ otherwise. When $c_1 = 1$ and $c_2 = 0$ or $c_1 = 0$ and $c_2 = 1$, the event is whether the Gini index in one survey is greater than c_3. When $c_1 = -1$ and $c_2 = 1$, the event is whether the change in the Gini index from period 1 to period 2 is greater than c_3, and so on.

The algorithm resembles the semiparametric bootstrap procedure proposed in Cowell and Flachaire (2007) and further explored in Davidson and Flachaire (2007), which provides valid refinements to the computations of confidence intervals and inference for inequality measures under the presence of extreme values. The basic idea is to replace the sample's upper tail with an equally sized sample of synthetic observations drawn from the Pareto distribution, while simultaneously resampling with replacement the rest of the distribution.

The algorithm in detail for both $t = \{1, 2\}$ is as follows:

1. Choose: $P \in (0, 1)$, the proportion of top income households; S, the top income share; N, the size of vector $\mathbf{y}$ used to compute the corrected Gini index; and B, the number of replications in the frequency crude simulator. All these quantities can be specific to each survey but we set them equal for brevity and clarity. Note that, given P, the minimum income of the top income group in survey t can be computed as $Y_{Pt} = \mathbf{Y}_t((1 - P)N_t)$, which is used to compute ρ_t and the tail index α_t from Eq. 7.
2. To generate a vector of pseudo data $\mathbf{y}_t$ of size N from for data vector $\mathbf{Y}_t$, draw N random numbers from the standardized uniform distribution $\{u_1, u_2, \ldots, u_N\}$. Then, for $i = 1, 2, \ldots, N$:
   (a) If $u_i \leq 1 - P$, the i-th observation of the t-th pseudo sample is the observation with rank $N_t u_i$ from $\mathbf{Y}_t$. That is to say, $\mathbf{y}_t(i) = \mathbf{Y}_t(N_t u_i)$.
   (b) If $u_i > 1 - P$, the observation is drawn from a Pareto distribution with tail index α_t and minimum observation Y_{Pt}. At this stage u_i is uniformly distributed on $(1 - P, 1)$ and so $w_i = (u_i - (1 - P))/P$ is uniformly distributed on $(0, 1)$. Thus, $\mathbf{y}_t(i) = Y_{Pt}(w_i)^{-1/\alpha_t}$.[14]
3. Compute the Gini indices and other statistics from the pseudo data.
4. Repeat steps 2 and 3 a large number of times, B.

4 Empirical and simulation analysis

In this section we present our empirical exploration using Peruvian data. In particular, we illustrate various possibilities to calibrate the top income share, in order to obtain probabilistic results about the magnitudes of the likely corrections to the Gini indices, and about their behavior through time.

[14] A computationally more efficient alternative to step 2 when vectorization of the operations is possible is to draw a random number n from a binomial distribution with N_t number of trials and probability of success P. Then, draw n observations from the Pareto distribution, and resample with replacement $N_t - n$ observations from the non-top income group.

4.1 Data and descriptive statistics

We use the publicly available household surveys ENAHO from 2004 to 2018.[15] The ENAHO is the annual nationally representative survey maintained by the Peruvian Statistics Bureau (INEI) and used to compute the official poverty rates. Also, it is used for international comparative studies (see Ferreira et al. 2013; Cord et al. 2015; Stampini et al. 2016) and for general research on the Peruvian living conditions (see Yamada et al. 2012; Herrera 2017). The scope and complexity of the survey has evolved over time with the number of households steadily increasing from about 19,000 in 2004 to more than 37,000 in 2018. Our definition of income is a summary measure of income before taxes computed directly by the INEI for official purposes, which we then divide by number of household members to obtain an approximation of individual level income. We call this definition *per-capita income before taxes* throughout the text. The results below are remarkably robust to the definition of income used and to the inclusion of survey weights.[16]

Figure 3 shows the evolution of various statistics related to income inequality and computed directly from ENAHO. Panel (a) shows the scaled mean-to-top statistic ρ defined in Eq. 6, and their confidence intervals, for $P = 1\%$ and $P = 2\%$. Since ρ is computed from the bottom $(1 - P)\%$ incomes, it is by construction independent of the behavior, or the mismeasurement, of the upper tail. For $P = 1\%$, the ratio increases steadily from 13 in 2004 to 16 in 2018, a 23 percent increase. This change implies that in annual terms the mean income μ^* has grown approximately 1.5 percent more than the threshold Y_P. The rate decreases, but remains significant, to around 1 percent for $P = 2\%$ (i.e., when Y_P gets closer to μ^*). These dynamics have been studied in Genoni and Salazar (2015), that describe how growth in Peru became inclusive and promoted shared prosperity, as income grew faster for low-skilled workers (Paz and Urrutia 2015) and initially disadvantaged groups (Winkelried and Torres 2019). Then, the reduction of the Gini index of the bottom 99%, G^* shown in Panel (d), from 48 to 44, comes at no surprise.

Regarding the upper part of the distribution. Panel (b) shows estimates of the tail index α.[17] Curiously, the point estimates are within the range (2, 3) documented by Ibragimov and Ibragimov (2018) for developed countries. It is interesting to note that α increases through time, from 2 to 3, significantly: the confidence intervals at the end of the period do not overlap with those at the beginning, so the null hypothesis that α remains unchanged will

[15] The ENAHO is available since 1997, but a major methodological revision to enhance the homogeneity and comparability among waves took place in 2004 (see Winkelried and Torres 2019). The ENAHO is also the primary source of the `WDI` statistics in Section 2.1. Our descriptive statistics are, of course, close to the `WDI`'s, but may not identical due to differences in the weighting scheme used.

[16] In ENAHO, the code for the total household income before taxes is `inghog1d` and for the number of household members is `mieperho`. We repeated our entire empirical analysis with six different definitions of income, using labor income and income after taxes, with qualitatively similar results. These are presented in Online Supplement A. On the other hand, in Online Supplement B we generalize the sampling algorithm to include survey weights, show weighted descriptive statistics and further simulations.

[17] The estimators of α exploit that the top P observations do belong to the Pareto tail of the true income distribution (see Ruiz and Woloszko 2016; Ibragimov and Ibragimov 2018). Charpentier and Flachaire (2019) review how the 'Rank-Size' equation derived from Eq. 1, $\log(i) = c - \alpha \log(Y_{(i)})$ where $Y_{(i)} > Y_P$ is the income in the i-th position of the upper group, can be used to estimate α. The estimation faces a bias-variance trade-off, with the choice of P being crucial, as the bias is reduced at the cost of a higher variance as $P \to 0$ (see Eckerstorfer et al. 2016). Gabaix and Ibragimov (2011) offer a simple but important improvement that we adopt. They show that estimating the equation $\log(i - 1/2) = c - \alpha \log(Y_{(i)})$ instead delivers an almost unbiased estimator that is asymptotically normal with a standard deviation of $\alpha\sqrt{2}$. The standard errors using the bootstrap procedure of Cowell and Flachaire (2007) were very closed to this asymptotic approximation.

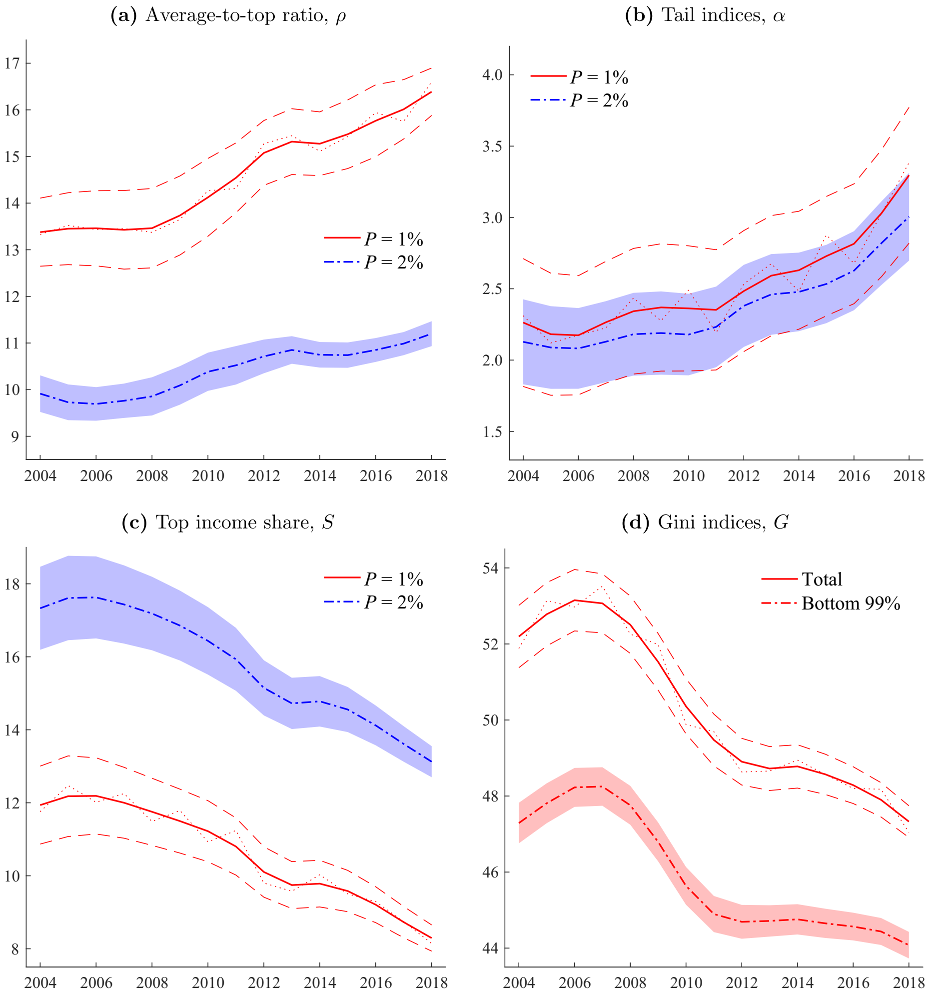

Fig. 3 Income inequality in the Peruvian household surveys, 2004 to 2018 **Source:** ENAHO, rounds from 2004 to 2018. Own elaboration. **Notes:** Computations using per capita household income before taxes. The 95% confidence intervals in panels (a), (c) and (d) were computed using a standard iid bootstrap with 5,000 replications. The tail indices in panel (b) are 'Rank-1/2' estimates (Gabaix and Ibragimov 2011) with asymptotic 95% confidence intervals. To ease visualization, all series were slightly smoothed with a 3-point centered moving average

be rejected. Thus, in the absence of misreporting of top incomes, the survey evidence points to a sustained reduction of inequality within the top income group, see Eq. 2.

Panel (c) shows that the (uncorrected) top income share computed directly from the household survey decreases gradually and sustainable, from 12 percent in 2004 to about 8 percent in 2018 for $P = 1\%$, or from 17 to 14 percent for $P = 2\%$.

Finally, panel (d) shows the (uncorrected) Gini index for the whole population: after a slight increase from 52 points in 2004 to 53 points in 2007, the index experiences a sustained 6-point decrease to 47 points in 2018, i.e., about 0.5 points a year. All the previous findings are consistent with this reduction: the decrease of inequality within the bottom $(1 - P)\%$

group shown in Panel (a), the decrease of inequality within the top $P\%$ group shown in Panel (b), and the decrease in the between component of inequality S shown in Panel (c). The vertical differences between the total Gini index and that of the bottom group (between 4 and 6 points) correspond neatly to the expression $G - G^* = (1 - G^*)S$ discussed in Section 3.2.

Next, we take a skeptical stance at the survey estimations on α and S, and show how different conjectures about these quantities affect our assessment on income inequality in Peru. It is important to mention our simulations below are admittedly stringent. All our simulations below use a pseudosample size $N = 10,000$, $B = 5,000$ replications, and focus on the case where $P = 1\%$.

4.2 Bounds for the top income share

Unlike the decreasing path of S in Fig. 3c, the corrected top income share reported in Flores et al. (2020) for various Latin American countries appear to be stable over similar time spans. Thus, when simulating inequality measures across time we will fix the same top income share for all years in our sample period.[18] In a first exercise we set the value of $S = 15$. The results are in Fig. 4a, which shows the mean and the 95% bounds of the simulated corrected Gini indices. There is some initial overlapping with the confidence intervals of the ENAHO Gini index, until 2010, as the conjecture $S = 15$ is relatively close to the initial sample estimate of S. After 2010, however, the corrected Gini index decreases slowly. This reveals that an important driver of the much faster reduction of the ENAHO Gini index is the decrease in the between group inequality, that is held approximately constant (since $S = 15$ all over the period) in the simulations.

In Fig. 4b we repeat the simulations but with $S = 20$. As this value is further from the sample statistics, the corrected Gini index also appears further from the ENAHO Gini index. Also, given the bottom 99% part of the distribution, $S = 20$ can be achieved only with much heavier tails than suggested by the sample, with implied values of α, as determined in Eq. 7, lower than 1.5. Therefore, the simulations involving such heavy-tailed distribution display naturally a wider range of variation.

Whether $S = 20$ is a reasonable assumption or not is probably the most important matter of debate in our application, especially in the absence of additional information about the upper tail. However, there are at least four indications that $S = 20$ can be taken as an upper bound in the Peruvian case (with the lower bound as indicated by ENAHO). First, as mentioned, $S = 20$ renders implausibly low values of α not only well below the confidence intervals found with the survey data, but also below the estimates documented in the literature worldwide (see, *inter alia*, Gabaix and Ibragimov 2011; Ibragimov and Ibragimov 2018). Second, by the end of the sample, $S = 20$ implies a corrected top income share that is 2.5 times as large as the survey share. When available, a vast majority of cases have the corrected share between 1.5 and 2 times as large as the uncorrected one (see, *inter alia*, Flores et al. 2020), with only extreme cases such as Brazil (Morgan 2018) or countries in the Middle East (Alvaredo et al. 2019) reporting ratios in the order of 2.5 to 3.

Third, following Anand and Segal (2017) a top income share can be imputed as the prediction of a cross-country regression. In particular, we regress the logarithm of the top 1% shares from the `WID.World` database, for 33 countries and for the periods 2000/2003 to

[18] Since the sample top income share is decreasing through time, keeping S fixed implies corrections to the inequality indices are increasing. This results in a conservative assessment of the reduction in inequality in Peru.

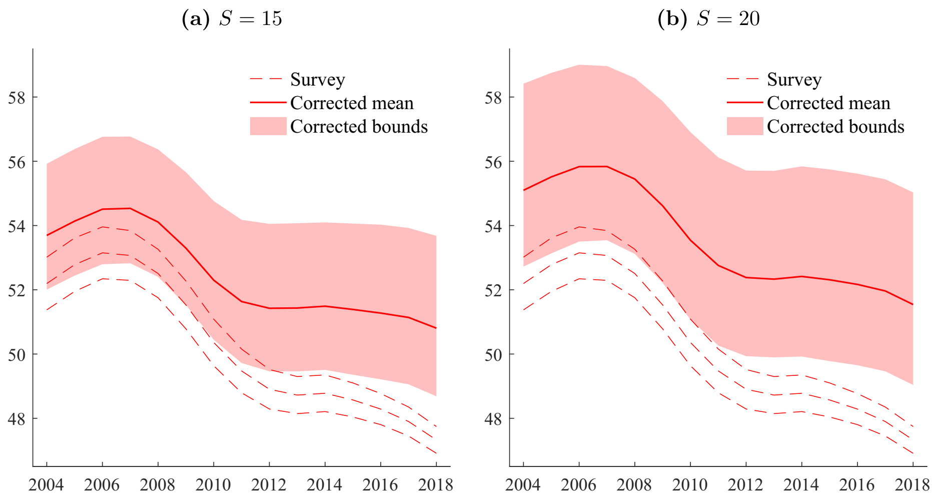

Fig. 4 Sample and corrected Gini indices, 2004 to 2018 **Source:** ENAHO, rounds from 2004 to 2018. Own elaboration. **Notes:** Sample and simulated Gini indices. The shaded areas represent a 95% interval of the corrected Gini indices and the continuous line is the mean across 5,000 replications

2015/2018 (a selection is shown in Fig. 2), on the logarithm of the uncorrected survey-based Gini coefficients from the `WDI`. As a result we obtain $\ln(\hat{S}) = -5.39_{(0.41)} + 0.95_{(0.11)} \ln(\underline{G})$ (standard errors in parenthesis), that produces a point prediction for the Peruvian top 1% share of 17.3 with a 99% prediction interval of (15.7, 19.3).[19]

Finally, according to the 2019 Forbes' World Billionaires records, Peru has 2 billionaires whereas its neighbor Chile has 7. Furthermore, the wealth of the Peruvian billionaires is 38 percent of that of the 2 top Chilean billionaires, whereas the Peruvian per capita GDP is 44 percent of the Chilean per capita GDP. Even though qualitative and superficial, this analysis strongly suggests that the Peruvian top 1% income share is unlikely to be larger than the Chilean counterpart. A similar conclusion can be reached from a comparison with Colombia, whose top billionaire is almost twice as wealthy as the Peruvian billionaires. The corrected top 1% income shares in this neighboring countries is about 20 per cent (see Flores et al. 2020).

4.3 Corrections to the Gini indices

Given a calibration of S, the simulations approximate the joint probabilities of Gini coefficients computed under different assumptions. Figure 4 shows the mean and the percentiles of these distributions through time. Another use of the simulations is to assess the magnitude of the corrections needed for the observed Gini index to be consistent with the data and various conjectures about S.

To elaborate, at a given year we can compute $\underline{G}$, the lower bound of the Gini index based on survey data and associated to the share $\underline{S}$, and a corrected version G with $S > \underline{S}$. The realizations of the random variables $\underline{G}$ and G are obtained using the same underlying draws $\{u_1, u_2, \ldots, u_N\}$ but two different tail indices: the sample estimate for $\underline{G}$, see Fig. 3d, and

[19] Further specifications, results and sensitivity analysis are available on request.

the one implied by Eq. 7 for G. Then, the difference $G - \underline{G}$ captures the magnitude of the biases contained in the lower bound (i.e., the size of the correction).

Figure 5 shows $\Pr(G - \underline{G} > c)$ for various values of c and selected years. Our previous finding that sizable corrections are more likely by the end of the sample period is confirmed in this analysis. For a given S, the probability curves shift upwards for all values of c as we move from 2004 to 2018. Alternatively, for a given value of c the corresponding curve shifts westwards with time, meaning that the corrections of size c are more likely with lower values of S, closer to the sample counterpart $\underline{S}$.

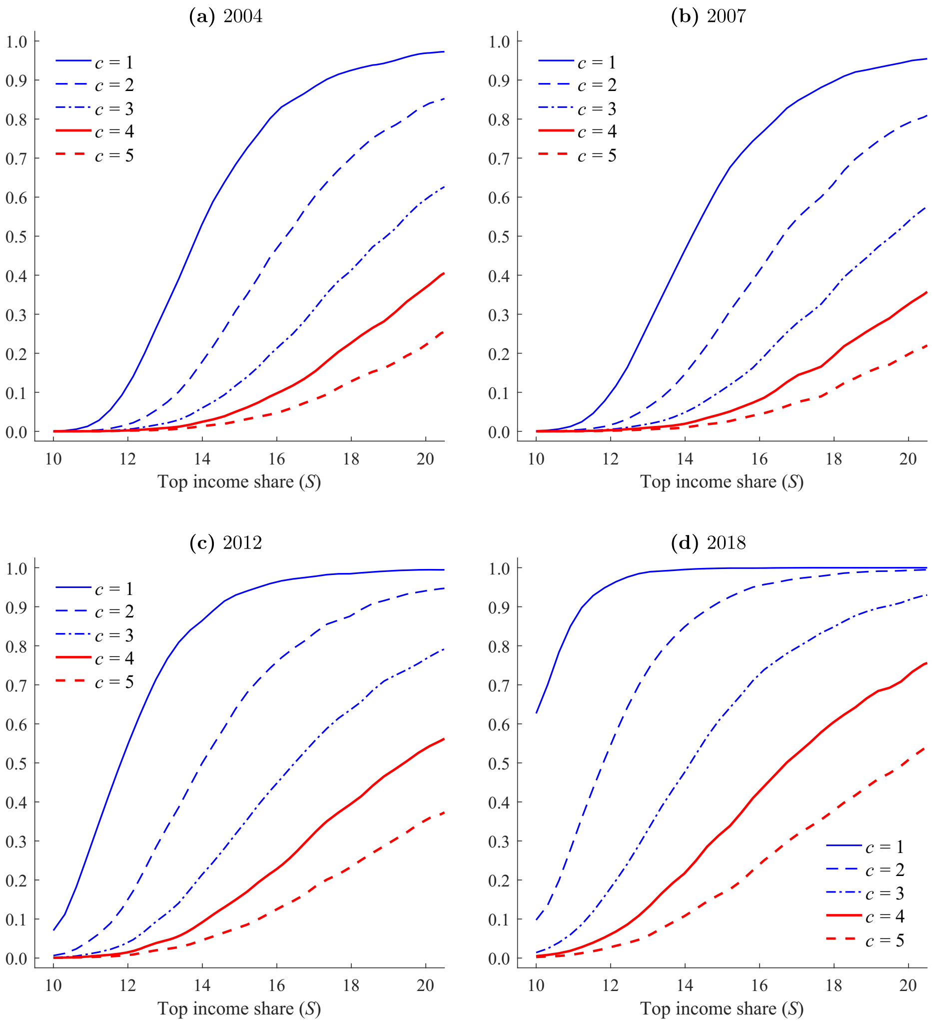

Fig. 5 Probabilities of correction in the Gini index by year, 2004 to 2018 **Source:** ENAHO, rounds from 2004 to 2018. Own elaboration. **Notes:** The graphs show $\Pr(G - \underline{G} > c)$ as a function of S. In words, this is the simulated probability of a correction in the Gini index of at least c points. The correction is the difference between the corrected Gini index simulated for an income share $S \in (10, 20)$, shown in the horizontal axes, and the ENAHO Gini index. All simulations use 5,000 replications

In order to infer a plausible upper bound for the bias in the Gini index from the abundant output of Figure 5, the researcher should choose a probability threshold π to define likely events. In other word, c is a likely bound if $\Pr(G - \underline{G} > c) \geq \pi$. For instance, if $\pi = 0.90$ and bearing in mind that a corrected top income share above 20 percent seems unlikely (so we focus on the $S \leq 20$ regions), we find that a correction of $c = 1$ point is quite likely in 2004 and 2007, as it is $c = 2$ points in 2012 and, to an extent, $c = 3$ points in 2018. Even though these choices of π and S are arbitrary, we do obtain the strong indication that $c > 3$ seems unlikely at any moment in the sample period.

In the Latin America studies that computed corrected Gini indices, $G - \underline{G}$ ranges normally from 1 to 5 points (Díaz-Bazán 2015; Higgins et al. 2018; Flores et al. 2020), but can be as high as 10 points (Alvaredo 2011). Thus, the plausible corrections for the Gini index in Peru are on par with the adjustments made for other Latin American countries.

4.4 Decrease through time

Is the observed reduction in inequality robust to the mismeasurement of the upper tail of the income distribution? Put differently, has the corrected Gini index experienced a similar decrease than the uncorrected index? The mean of the corrected Gini displays a downward trend in Fig. 3, even when S is as large as $S = 20$. Now we focus on the two points in time that mark the period when the ENAHO Gini index decreases steadily: 2007 and 2018. The qualitative conclusions using different windows are similar.

Figure 6 shows $\Pr(G_{2007} - G_{2018} > c)$ for various values of c, the amount of the reduction in the corrected index, and S. As before, these probabilities are decreasing in c for a given S, see panel (a). They are also decreasing in S only when c is less than 4, see panel (b). The probability that the corrected index decreased by at least $c = 2$ points is very high

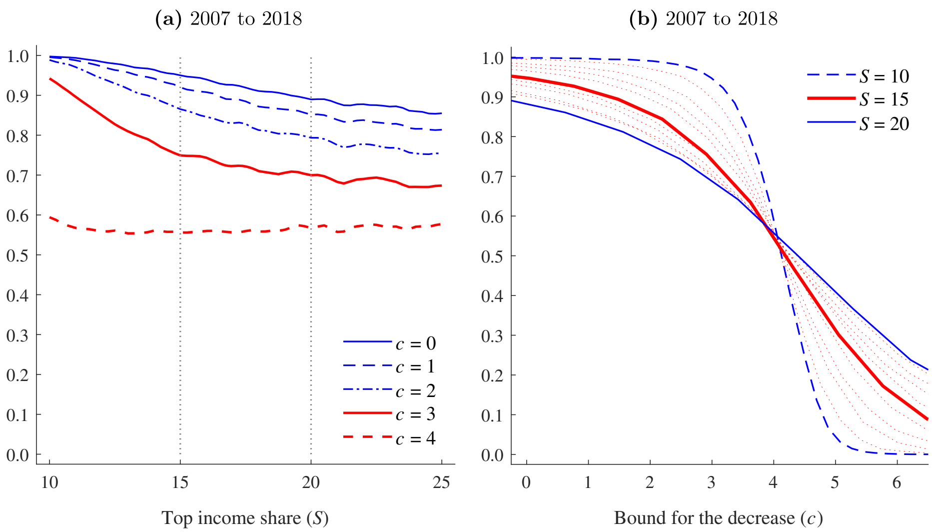

Fig. 6 Probabilities of change in the Gini index, 2007 to 2018 **Source:** ENAHO, rounds from 2004 to 2018. Own elaboration. **Notes:** The graphs show $\Pr(G_{2007} - G_{2018} > c)$. In words, the simulated probability of a decrease through time of the Gini index of at least c points. In panel (a) the probability is depicted as a function of S for selected values of c; in panel (b), as a function of c for selected values of S. All simulations use 5,000 replications

regardless of the value of S. It remains high for $c = 3$ points for values of S that, at most, double the survey figure, such as $S = 15$; for values closer to our suggested upper bound of $S = 20$, the probability decreases sharply to $\Pr(G_{2007} - G_{2018} > 4) \simeq 0.6$.

The actual difference observed in the survey dataset between the 2007 and 2018 Gini indices, and also the difference observed the simulated Gini series in Fig. 4 is about 4 points. Thus, that the probability of decrease with bound $c = 4$ is close to 0.5 reflects this stylized fact. For $c > 4$, probabilities sharply decrease in S as the statistics from the simulated datasets quickly reject a difference higher than that observed in the raw datasets. The probabilities for simulations with higher top income shares (e.g. $S = 20$ in panel b) do not decrease as quickly due to the higher variance of Gini estimates that we already documented in Fig. 4. Overall, we take the conservative stance that values $c > 4$ are unlikely.

We conclude that even when we introduce large corrections to the sequence of Gini indices, the sequence remains decreasing. The decrease in the corrected index is not as large as the 6-point reduction of the uncorrected index. These results are in line with the findings in Section 4.3 that, given a fixed S, the upwards adjustment to the Gini indices are increasing in time (larger by the end of the sample). Yet, there is a strong indication of a reduction of $c \in (3, 4)$.

5 Closing remarks

The reduction of inequality in Latin America during the 2000s and 2010s is a cornerstone in the narrative of positive economic development in the region, as represented by the Peruvian experience we study. We enquiry whether valid concerns about underestimation of top income shares in household surveys undermine such a favorable trend in inequality. We achieve this by developing a simple simulation method that does not require additional information, such as tax records. The method maps from the data and the calibration of a single and easy-to-interpret parameter, the top income share, to probability statements and ranges of values – rather than point estimates – of the inequality indices.

Using stringent calibrations for the unobserved top income share, which are in line with estimates from countries regarded as less egalitarian, we find that in Peru the bias in the Gini index is at most 3 points, so by the end of our sample the index should be within the (50, 53) range rather than the (47, 49) range reported in the surveys. Likewise, the uncorrected Gini index decreased in about 6 points from 2007 to 2018, whereas our analysis indicates that the reduction is more likely to be around 3 points. Despite the upward adjustments, the study still shows a strong case for a sustained reduction in inequality in Peru.

In our approach, gaps between the observable income distribution in surveys and the unknown underlying distribution are filled in with calibrations that, even though may be educated, are essentially untested prior information. We consider that a fruitful topic for future research is to give a fully-fledged Bayesian treatment to this problem, not only to be able to obtain posterior inferences about the unobserved S but also as a means to combine in an statistically correct manner the information of these calibrations with other sources such as tax data whenever it is available. The specification for misreporting and non-response proposed in Blanchet et al. (2018) also constitutes an interesting direction for us to expand our algorithm. Finally, a natural empirical extension would be to apply our method to a harmonized income database, for example for the Latin American region, to revise regional trends more thoroughly.

6 Competing interests

The authors declare to have no competing interests or other interests that might be perceived to influence the elaboration and interpretation of this article.

7 Data availability

All data used in this article and its online supplement are publicly available at the website of the Peruvian Statistics Bureau, INEI: http://iinei.inei.gob.pe/microdatos/. In particular, data come from the summary files (module 34, "Sumaria") of all ENAHO waves. The codenames of the variables used (such as `inghog1d` and `mieperho`) can be found in the text and the supplement.

References

Aaberge, R., Atkinson, A.B., Modalsli, J.: Estimating long-run income inequality from mixed tabular data: Empirical evidence from Norway, 1875–2017. J. Public Econ. **187**, 104196 (2020)
Alfons, A., Templ, M., Filzmoser, P.: Robust estimation of economic indicators from survey samples based on Pareto tail modelling. J. R. Stat. Soc. Ser. C (Appl. Stat.) **62**(2), 271–286 (2013)
Alvaredo, F.: The rich in Argentina over the twentieth century, 1932-2004. In: Atkinson, A.B., Piketty, T. (eds.) Top Incomes: A Global Perspective, chapter 6, pp. 253–298. Oxford University Press, Oxford (2010)
Alvaredo, F.: A note on the relationship between top income shares and the Gini coefficient. Econ. Lett. **110**(3), 274–277 (2011)
Alvaredo, F., Assouad, L., Piketty, T.: Measuring inequality in the Middle East 1990-2016: The world's most unequal region? Rev. Income Wealth **65**(4), 685–711 (2019)
Alvaredo, F., Atkinson, A.B., Blanchet, T., Chancel, L., Bauluz, L., Fisher-Post, M., Flores, I., Garbinti, B., Goupille-Lebret, J., Martínez-Toledano, C., et al.: Distributional national accounts guidelines, methods and concepts used in the world inequality database. Technical report (2020)
Alvaredo, F., Londoño Vélez, J.: High Incomes and Personal Taxation in a Developing Economy: Colombia 1993-2010 CEQ Working Paper 12. CEQ Institute, Tulane University (2013)
Amarante, V.: Income inequality in Latin America: A factor component analysis. Rev. Income Wealth **62**, S4–S21 (2016)
Anand, S., Segal, P.: Who are the global top 1%? World Dev. **95**, 111–126 (2017)
Assouad, L., Chancel, L., Morgan, M.: Extreme inequality: Evidence from Brazil, India, the Middle East, and South Africa. AEA Papers Proc. **108**, 119–23 (2018)
Atkinson, A.B.: Top incomes in the UK over the 20th century. J. R. Stat. Soc. Ser. A (Stat. Soc.) **168**(2), 325–343 (2005)
Atkinson, A.B.: Measuring top incomes: Methodological issues. In: Atkinson, A.B., Piketty, T. (eds.) Top Incomes Over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries, chapter 1, pp. 18–42. Oxford University Press, Oxford (2007)
Atkinson, A.B., Piketty, T.: Top Incomes: a Global Perspective. Oxford University Press, Oxford (2010)
Atkinson, A.B., Piketty, T., Saez, E.: Top incomes in the long run of history. J. Econ. Lit. **49**(1), 3–71 (2011)
Auten, G., Splinter, D.: Top 1 percent income shares: Comparing estimates using tax data. AEA Papers Proc. **109**, 307–11 (2019)
Balkema, A.A., De Haan, L.: Residual life time at great age. Ann. Probab. **2**(5), 792–804 (1974)
Bee, M.: Estimation of the Lognormal-Pareto distribution using probability weighted moments and maximum likelihood. Commun. Stat. - Simul. Comput. **44**(8), 2040–2060 (2015)
Blanchet, T., Chancel, L., Gethin, A.: How Unequal Is Europe? Evidence from Distributional National Accounts, 1980-2017. WID.World Working Paper 2019/06, World Inequality Lab (2019)
Blanchet, T., Flores, I., Morgan, M.: The weight of the rich: Improving surveys using tax data. WID.World Working Paper 2018/12, World Inequality Lab (2018)
Bourguignon, F.: Appraising income inequality databases in Latin America. J. Econ. Inequal. **13**(4), 557–578 (2015)

Burdín, G., De Rosa, M., Vigorito, A., Vilá, J.: Was falling inequality in all Latin American countries a data-driven illusion? income distribution and mobility patterns in Uruguay 2009-2016. Discussion Paper 13070, IZA - Institute of Labor Economics (2020)

Burkhauser, R., Hahn, M., Wilkins, R.: Measuring top incomes using tax record data: A cautionary tale from Australia. J. Econ. Inequal. **13**(2), 181–205 (2015)

Burkhauser, R.V., Feng, S., Jenkins, S.P., Larrimore, J.: Recent trends in top income shares in the united states: Reconciling estimates from March CPS and IRS tax return data. Rev. Econ. Stat. **94**(2), 371–388 (2012)

Burkhauser, R.V., Hérault, N., Jenkins, S.P., Wilkins, R.: What has been happening to UK income inequality since the mid-1990s? answers from reconciled and combined household survey and tax return data. ISER Working Paper 2016-03, Institute for Social and Economic Research (2016)

Chancel, L., Cogneau, D., Gethin, A., Myczkowski, A.: How Large Are African Inequalities? Towards Distributional National Accounts in Africa, 1990-2017. WID.World Working Paper 2019/13, World Inequality Lab. (2019)

Chancel, L., Piketty, T.: Indian Income Inequality, 1922-2015: From British Raj to Billionaire Raj? Rev. Income Wealth **65**, S33–S62 (2019)

Charpentier, A., Flachaire, E.: Pareto Models for Top Incomes. Technical report (2019)

Clementi, F., Gallegati, M.: Pareto's law of income distribution: Evidence for Germany, the United Kingdom, and the United States. In: Chatterjee, A., Yarlagadda, S., Chakrabarti, B.K. (eds.) Econophysics of Wealth Distributions, chapter 1, pp. 3–14. Springer, Milano (2005)

Cord, L., Genoni, M.E., Rodríguez-Castelán, C.: Overview. In: Cord, L., Genoni, M.E., Rodríguez-Castelán, C. (eds.) Shared prosperity and Poverty Eradication in Latin America and the Caribbean, chapter 1, pp. 1–51. World Bank, Washington (2015)

Cowell, F.A., Flachaire, E.: Income distribution and inequality measurement: The problem of extreme values. J. Econ. **141**(2), 1044–1072 (2007)

Davidson, R., Flachaire, E.: Asymptotic and bootstrap inference for inequality and poverty measures. J. Econ. **141**(1), 141–166 (2007)

De Rosa, M., Flores, I., Morgan, M.: Inequality in Latin America revisited: Insights from Distributional National Accounts. Technical Note 2020/02, World Inequality Lab. (2020)

Deaton, A.: Measuring poverty in a growing world (or measuring growth in a poor world). Rev. Econ. Stat. **87**(1), 1–19 (2005)

Díaz-Bazán, T.: Measuring Inequality from Top to Bottom. World Bank Policy Research Paper 7237, World Bank (2015)

Eckerstorfer, P., Halak, J., Kapeller, J., Schütz, B., Springholz, F., Wildauer, R.: Correcting for the missing rich: an application to wealth survey data. Rev. Income Wealth **62**(4), 605–627 (2016)

Ferreira, F.H.G., Messina, J., Rigolini, J., López-Calva, L.F., Lugo, M.A., Vakis, R.: Economic Mobility and the Rise of the Latin American Middle Class. The World Bank, Washington (2013)

Flores, I., Sanhueza, C., Atria, J., Mayer, R.: Top incomes in Chile: A historical perspective on income inequality, 1964-2017. Rev. Income Wealth **66**(4), 850–874 (2020)

Gabaix, X., Ibragimov, R.: Rank$-1/2$: A simple way to improve the OLS estimation of tail exponents. J. Bus Econ. Stat. **29**(1), 24–39 (2011)

Gasparini, L.: Different lives: Inequality in Latin America. In: De Ferranti, D., Perry, G.E., Ferreira, F.H., Walton, M. (eds.) Inequality in Latin America and the Caribbean: Breaking with History?, chapter 2, pp. 35–71. World Bank, Washington (2004)

Gasparini, L., Cruces, G., Tornarolli, L., Mejía, D.: Recent trends in income inequality in Latin America. Economía **11**(2), 147–201 (2011)

Genoni, M.E., Salazar, M.: Steering towards shared prosperity in Peru. In: Cord, L., Genoni, M.E., Rodríguez-Castelán, C. (eds.) Shared Prosperity and Poverty Eradication in Latin America and the Caribbean, chapter 8, pp. 269–301. World Bank, Washington (2015)

Hajargasht, G., Griffiths, W.E.: Pareto-lognormal distributions: inequality, poverty, and estimation from grouped income data. Econ. Model. **33**, 593–604 (2013)

Herrera, J.: Poverty and economic inequalities in Peru during the boom in growth: 2004-14. In: Carbonnier, G., Campodónico, H., Vázquez, S.T. (eds.) Alternative Pathways to Sustainable Development: Lessons from Latin America, chapter 8, pp. 138–173. Brill Nijhoff (2017)

Higgins, S., Lustig, N., Vigorito, A.: The rich underreport their income: Assessing bias in inequality estimates and correction methods using linked survey and tax data. ECINEQ WP 2018-475, ECINEQ (2018)

Ibragimov, M., Ibragimov, R.: Heavy tails and upper-tail inequality: The case of Russia. Empir. Econ. **54**(2), 823–837 (2018)

Jenkins, S.P.: Pareto models, top incomes and recent trends in UK income inequality. Economica **84**(334), 261–289 (2017)

Kim, N.-N.: Top incomes in korea: update, 1933-2016. Korean Econ. Forum **11**(1), 1–21 (2018)

Lakner, C., Milanovic, B.: Global income distribution: from the fall of the Berlin Wall to the Great Recession. World Bank Econ. Rev. **30**(2), 203–232 (2016)

Lustig, N.: The 'Missing Rich' in Household Surveys: Causes and Correction Approaches. Working Paper 75 Commitment to Equity (CEQ) Institute. Tulane University (2019)

Lustig, N., López-Calva, L.F., Ortiz-Juarez, E.: Declining inequality in Latin America in the 2000s: The cases of Argentina, Brazil, and Mexico. World Dev. **44**, 129–141 (2013)

Morgan, M.: Income inequality, growth and elite taxation in Brazil: New evidence combining survey and fiscal data, 2001-2015. Working Paper 165, International Policy Centre for Inclusive Growth (2018)

Novokmet, F., Piketty, T., Yang, L., Zucman, G.: From communism to capitalism: Private versus public property and inequality in China and russia. In: AEA Papers and Proceedings, vol. 108, pp 109–13 (2018)

Paz, P., Urrutia, C.: Economic growth and wage stagnation in Peru: 1998-2012. Rev. Dev. Econ. **19**(2), 328–345 (2015)

Piketty, T.: Income inequality in France, 1901-1998. J. Polit. Econ. **111**(5), 1004–1042 (2003)

Piketty, T., Saez, E.: Income inequality in the United States, 1913-1998. Q. J. Econ. **118**(1), 1–41 (2003)

Piketty, T., Saez, E., Zucman, G.: Distributional national accounts: Methods and estimates for the United States. Q. J. Econ. **133**(2), 553–609 (2018)

Piketty, T., Yang, L., Zucman, G.: Capital accumulation, private property, and rising inequality in China, 1978-2015. Amer. Econ. Rev. **109**(7), 2469–96 (2019)

Reed, W.J.: The Pareto law of incomes: An explanation and an extension. Physica A: Stat. Mech. Appl. **319**, 469–486 (2003)

Reed, W.J., Jorgensen, M.: The double Pareto-lognormal distribution—a new parametric model for size distributions. Commun. Stat. - Theory Methods **33**(8), 1733–1753 (2004)

Ruiz, N., Woloszko, N.: What do household surveys suggest about the top 1% incomes and inequality in OECD countries? OECD Economics Department Working Paper 1265, OECD (2016)

Saez, E., Zucman, G.: Trends in US income and wealth inequality: Revising after the revisionists. Working Paper 27921, National Bureau of Economic Research (2020)

Smith, M., Yagan, D., Zidar, O., Zwick, E.: Capitalists in the twenty-first century. Q. J. Econ. **134**(4), 1675–1745 (2019)

Smith, M., Zidar, O.M., Zwick, E.: Top wealth in America: New estimates and implications for taxing the rich. Working Paper 29374, National Bureau of Economic Research (2021)

Stampini, M., Robles, M., Sáenz, M., Ibarrarán, P., Medellín, N.: Poverty, vulnerability, and the middle class in Latin America. Latin Amer. Econ. Rev. 25(1):Article 4 (2016)

Székely, M., Mendoza, P.: Declining inequality in Latin America: Structural shift or temporary phenomenon? Oxf. Dev. Stud. **45**(2), 204–221 (2016)

Vermeulen, P.: How fat is the top tail of the wealth distribution? Rev. Income Wealth **64**(2), 357–387 (2018)

Winkelried, D., Torres, J.: Economic mobility along the business cycle. the case of Peru. Appl. Econ. **51**(18), 1894–1906 (2019)

Yamada, G., Castro, J.F., Bacigalupo, J.: Desigualdad monetaria en un contexto de rápido crecimiento económico: El caso reciente del Perú. Rev. Estud. Econ. **24**, 65–77 (2012)