



Munich Personal RePEc Archive

Declining inequality in Latin America? Robustness checks for Peru

Winkelried, Diego and Escobar, Bruno

Universidad del Pacifico

16 December 2020

Online at <https://mpra.ub.uni-muenchen.de/106566/>
MPRA Paper No. 106566, posted 12 Mar 2021 21:36 UTC

Declining inequality in Latin America? Robustness checks for Peru*

Diego Winkelried[†]
winkelried_dm@up.edu.pe

Bruno Escobar
b.escobarizquierdo@up.edu.pe

School of Economics and Finance, Universidad del Pacífico (Lima, Peru)

This version: December 16, 2020

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., Chile). We find that even though the point estimates of the Gini index are biased indeed, the corrected indices still decrease in time.

JEL Classification : D63, O54, F63, N36

Keywords : Top income share, income inequality, Latin America.

*We are indebted to Gustavo Yamada for insightful discussion. We would also like to thank Nelson Ramírez, Marco Terrones, Javier Torres and seminar participants at the Research Center of Universidad del Pacífico (CIUP) for useful comments. The financial support of the CIUP is gratefully acknowledged. We alone are responsible for the views expressed in this paper and for any remaining errors.

[†]Corresponding author. School of Economics and Finance, Universidad del Pacífico, Av. Salaverry 2020, Lima 11, Peru. Phone: (+511) 219 0100 ext. 2238.

Declining inequality in Latin America? Robustness checks for Peru

This version: December 16, 2020

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., Chile). We find that even though the point estimates of the Gini index are biased indeed, the corrected indices still decrease in time.

JEL Classification : D63, O54, F63, N36

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 (from 53 to 45 points) in that period, significantly more than in any other region in the world. These findings have fed an academic 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 understate the income of the richest households and, therefore, underestimate inequality ([Deaton, 2005](#); [Lustig, 2019](#)). A common practice to correct underreporting – 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., 2019](#)), Colombia ([Alvaredo and Londoño Vélez, 2013](#); [Díaz-Bazán, 2015](#)) and Uruguay ([Higgins et al., 2018](#); [Burdín et al., 2020](#)).

However, tax data are not available for all countries or are costly to update. Also, corrections made to inequality indices are not necessarily trustworthy, while being subject to some degree of uncertainty; hence, the literature calls for a focus on ranges rather than point estimates ([Burkhauser et al., 2015](#); [Jenkins, 2017](#); [Auten and Splinter, 2019](#); [Lustig, 2019](#); [Aaberge et al., 2020](#)). 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 without the need of extra data. Borrowing ideas from the semiparametric bootstrap scheme advanced in [Cowell and Flachaire \(2007\)](#) and [Davidson and Flachaire \(2007\)](#), the idea 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 without top income underreporting which can help to track the evolution of corrected, or ‘true’, inequality measures and its uncertainty.¹

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 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.

¹ 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.

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 and Chile, that is about 20 percent. By using 20 percent as the top income share of the unobserved true income distribution, we obtain the joint distributions of adjusted Gini indices and 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 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.²

Figure 1(a) 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 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.³ 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 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

² 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.

³ The high volatility of the African data is due to the diversity of inequality patterns thoroughly studied in [Chancel et al. \(2019\)](#).

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 1(b) 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, Figure 1(c) 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, Figure 1(d) 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 living standards surveys tend to underreport the household incomes, especially of those at the top of the distribution (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, to refuse answering questions related to income declaration, or to consistently underreport their income. These factors cause sparseness (i.e., the lack of density mass) of income in the upper tail and, in severe cases, right truncation (Jenkins, 2017).⁴ As a result, inequality estimates based on households surveys are biased downwards, and a growing body of research has been developed to correct them (see, *inter alia* Atkinson et al., 2011; Ruiz and Woloszko, 2016).

A first branch seeks to improve the data used in the assessment of inequality. 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

⁴ 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.

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 per capita income and per capita income from the national accounts, often per capita GDP, to the share of income of the unobserved upper tail. 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 sizeable and artificial changes in inequality estimates (see Deaton, 2005; Bourguignon, 2015).

A compromise of the aforementioned methods, the so-called the ‘Distributional National Accounts’ (DINA) popularized in Piketty et al. (2018), combines national accounts, survey and tax data to estimate a complete distribution of income that represents the whole population and that matches macroeconomic aggregates. Tax-based methods are the main tool to produce the results gathered in the World Inequality Database (WID.world), a publicly available collection of top incomes shares around the globe. The country selection is quite diverse: 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 American, the selection includes Argentina (Alvaredo, 2010, 2011), Brazil (Morgan, 2018; Assouad et al., 2018), Chile (Flores et al., 2019), Colombia (Alvaredo and Londoño Vélez, 2013; Díaz-Bazán, 2015) and Uruguay (Higgins et al., 2018; Burdín et al., 2020). Despite its popularity, tax-based and DINA methods are not free from critiques. As shown in Auten and Splinter (2019), they 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.

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.

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 with a similar 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.⁵

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 underreporting, 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.

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.

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 Gallegati, 2005](#); [Cowell and Flachaire, 2007](#); [Charpentier and Flachaire, 2019](#)) both as an exact model for the upper tail, or as a useful approximation.⁶

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

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

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 :

$$\begin{aligned} \text{Gini}(Y | Y \geq Y_P) &= \frac{1}{2\alpha - 1}, \quad \text{E}(Y | Y \geq Y_P) = \left(\frac{\alpha}{\alpha - 1}\right) Y_P \\ \text{and } \text{var}(Y | Y \geq Y_P) &= \frac{\alpha}{\alpha - 2} \left(\frac{Y_P}{\alpha - 1}\right)^2. \end{aligned} \quad (2)$$

⁵ [Vermeulen \(2018\)](#) develops an innovative approach that involves both reweighing 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.

⁶ 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.

From (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, Alvarado (2011) shows that the Gini index in the whole population, G , can be written as:

$$G = P S G^{**} + (1 - P)(1 - S)G^* + (S - P). \quad (3)$$

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 (3) is the within-groups component, whereas the last term, that takes the very simple form $S - P$, is the between-groups component. When $P \rightarrow 0$ and S remains finite, G can be readily approximated by:

$$G = (1 - S)G^* + S = G^* + (1 - G^*)S, \quad (4)$$

The share S plays a key role in our analysis. Consider that, as usual, $G^* \in (0, 1)$. From (4), 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 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.⁷

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:⁸

$$S = \left[1 + \rho \left(1 - \frac{1}{\alpha} \right) \right]^{-1} \quad \text{where} \quad \rho = \left(\frac{1 - P}{P} \right) \frac{\mu^*}{Y_P}. \quad (5)$$

For $\alpha > 1$, this is a strictly decreasing function of α . When $\alpha \rightarrow 1$ from the right, then $S \rightarrow 1$. On the other hand, as $\alpha \rightarrow \infty$, S converges to the finite value $(1 + \rho)^{-1}$. It follows that the

⁷ 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 (4).

⁸ 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 $P \rightarrow 0$. In this case, (5) holds with $\rho = \mu^* / (PY_P)$.

Gini index of the whole population G is also a decreasing function of α , ranging from $G = 1$ when $\alpha \rightarrow 1$ from the right, to G close to G^* when $\alpha \rightarrow \infty$. The inverse function of (5) is:

$$\alpha = \left[1 - \frac{1}{\rho} \left(\frac{1}{S} - 1 \right) \right]^{-1}. \quad (6)$$

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 (6) 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.⁹

Any consistent estimator of the Gini index that use the data \mathbf{y} will converge in probability to the corrected G , as decomposed in (3). 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 $c_1 G_1 + c_2 G_2 > c_3$, for fixed c_1, c_2 and c_3 can be well approximated by a crude frequency counting:

$$\frac{1}{B} \sum_{b=1}^B \mathbb{I}(c_1 G_1^{(b)} + c_2 G_2^{(b)} > c_3) \xrightarrow{P} \Pr(c_1 G_1 + c_2 G_2 > c_3) \quad \text{as } B \rightarrow \infty, \quad (7)$$

where $\mathbb{I}(x)$ is the step function such that $\mathbb{I}(x) = 1$ if x is true, and $\mathbb{I}(x) = 0$ 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 (6).

⁹ An alternative route is 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.

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 $\{u_1, u_2, \dots, u_N\}$. Then, for $i = 1, 2, \dots, N$:
 - (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, $\mathbf{y}_t(i) = Y_{Pt}(w_i)^{-1/\alpha_t}$.¹⁰
3. Compute the Gini indexes 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.

4.1 Data and descriptive statistics

We use the publicly available household surveys ENAHO from 2004 to 2018.¹¹ 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 per capita income before taxes computed directly by the INEI for official purposes. The results below are remarkably robust to the definition of income used.¹²

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 (5), 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,

¹⁰ 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.

¹¹ 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 be identical due to differences in the weighting scheme used (our results are unweighed).

¹² In ENAHO, the code for the total household income before taxes is `inghog1d` and for the number of household members is `mieperho`. We repeat our entire empirical analysis with six different definitions, e.g. using labor income and income after taxes. The results were qualitatively similar and did not provide additional insights to those presented here. Thus, the results using these alternative definitions are detailed in an online supplement.

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 .¹³ Curiously, the point estimates are within the range $\in (2, 3)$ 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 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 equation (2).

Panel (c) shows that the (uncorrected) top income share computed directly from the household survey decreases gradually and sustainably, 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 an 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 Figure 3(c), the corrected top income share reported in [Flores et al. \(2019\)](#) 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

¹³ 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 (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 $P \rightarrow 0$ (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 \(2007\)](#) where very closed to this asymptotic approximation.

share for all years in our sample period.¹⁴ In a first exercise we set the value of $S = 15$. The results are in Figure 4(a), 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 Figure 4(b) 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 equation (6), 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., 2019), with only extreme cases such as Brazil (Assouad et al., 2018) or countries in the Middle East (Alvaredo and Londoño Vélez, 2013) 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 2015/2018 (a selection is shown in Figure 2), on the logarithm of the uncorrected survey-based Gini coefficients from the WDI. As a result we obtain $\ln(\hat{S}) = -6.04_{(0.47)} + 1.11_{(0.13)} \ln(\underline{G})$ (standard errors in parenthesis), that produces a point prediction for the Peruvian top 1% share of 16.9 with a 99% prediction interval of (15.0, 19.1).¹⁵

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., 2019).

¹⁴ Since the sample top income 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.

¹⁵ Further specifications, results and sensitivity analysis are available on request.

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 $\{u_1, u_2, \dots, u_N\}$ but two different tail indices: the sample estimate for \underline{G} , see Figure 3(d), and the one implied by equation (6) 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 sizeable 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} .

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 particular, a value of π to regard c as 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., 2019), but can be exceptionally 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 Figure 3, even when S is as large as $S = 20$. Now we focus on the two points in time that mark the period where 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 $c = 3$ points is very high regardless of the value of S . It remains high for $c = 4$ points but for values of S that doubles the survey figure, such as $S = 15$; for values closer to our suggested upper bound of $S = 20$, the probability decreases sharply to $\Pr(G_{2007} - G_{2018} > 4) \simeq 0.55$. The value of $c = 4$ can be

interpreted as a median decrease for most of the shares S included in the simulations, and 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 beginning 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 enquire whether valid concerns about undercoverage of top income household surveys undermine such a favorable trend in inequality, 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, 4). 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 believe 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.

References

- Aaberge, R., Atkinson, A. B., and Modalsli, J. (2020). Estimating long-run income inequality from mixed tabular data: Empirical evidence from Norway, 1875–2017. *Journal of Public Economics*, 187:104196.
- Alfons, A., Templ, M., and Filzmoser, P. (2013). Robust estimation of economic indicators from survey samples based on Pareto tail modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(2):271–286.
- Alvaredo, F. (2010). The rich in Argentina over the twentieth century, 1932–2004. In Atkinson, Anthony B. and Piketty, Thomas, editor, *Top Incomes: A Global Perspective*, chapter 6, pages 253–298. Oxford University Press, Oxford.
- Alvaredo, F. (2011). A note on the relationship between top income shares and the Gini coefficient. *Economics Letters*, 110(3):274–277.
- Alvaredo, F., Assouad, L., and Piketty, T. (2019). Measuring inequality in the Middle East 1990–2016: The world’s most unequal region? *Review of Income and Wealth*, 65(4):685–711.
- Alvaredo, F. and Londoño Vélez, J. (2013). High incomes and personal taxation in a developing economy: Colombia 1993–2010. CEQ Working Paper 12, CEQ Institute, Tulane University.
- Amarante, V. (2016). Income inequality in Latin America: A factor component analysis. *Review of Income and Wealth*, 62:S4–S21.
- Anand, S. and Segal, P. (2017). Who are the global top 1%? *World Development*, 95:111–126.
- Assouad, L., Chancel, L., and Morgan, M. (2018). Extreme inequality: Evidence from Brazil, India, the Middle East, and South Africa. *AEA Papers and Proceedings*, 108:119–23.
- Atkinson, A. B. (2005). Top incomes in the UK over the 20th century. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2):325–343.
- Atkinson, A. B. (2007). Measuring top incomes: Methodological issues. In Atkinson, A. B. and Piketty, T., editors, *Top Incomes Over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*, chapter 1, pages 18–42. Oxford University Press, Oxford.
- Atkinson, A. B. and Piketty, T. (2010). *Top Incomes: A Global Perspective*. Oxford University Press, Oxford.
- Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71.
- Auten, G. and Splinter, D. (2019). Top 1 percent income shares: Comparing estimates using tax data. *AEA Papers and Proceedings*, 109.
- Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.
- Bee, M. (2015). Estimation of the Lognormal-Pareto distribution using probability weighted moments and maximum likelihood. *Communications in Statistics - Simulation and Computation*, 44(8):2040–2060.
- Blanchet, T., Chancel, L., and Gethin, A. (2019). How Unequal Is Europe? Evidence from Distributional National Accounts, 1980–2017. WID.World Working Paper 2019/06, World Inequality Lab.
- Bourguignon, F. (2015). Appraising income inequality databases in Latin America. *Journal of Economic Inequality*, 13(4):557–578.

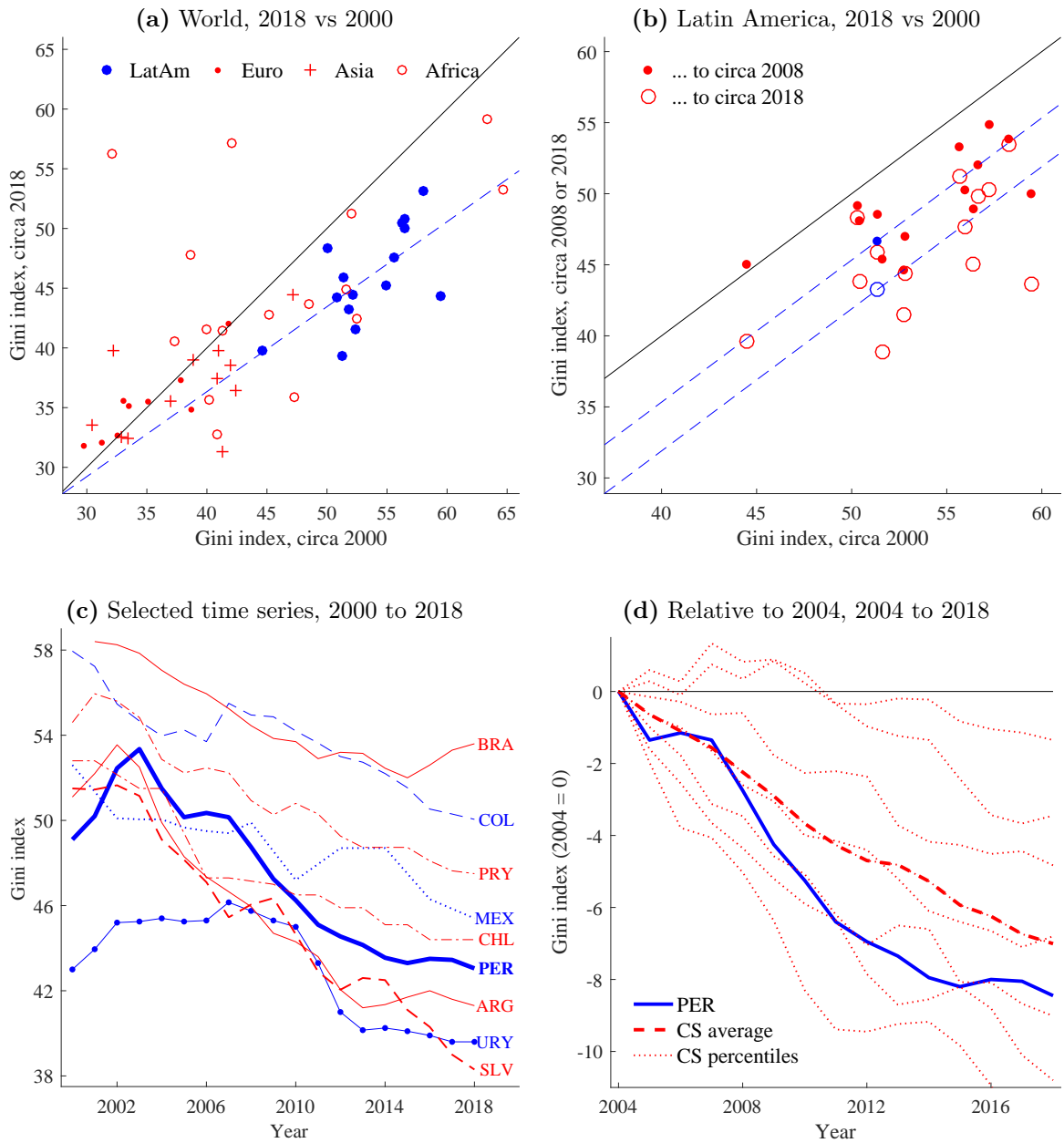
- Burdín, G., De Rosa, M., Vigorito, A., and Vilá, J. (2020). Was falling inequality in all Latin American countries a data-driven illusion? income distribution and mobility patterns in Uruguay 2009-2016. Discussion Paper 13070, IZA - Institute of Labor Economics.
- Burkhauser, R., Hahn, M., and Wilkins, R. (2015). Measuring top incomes using tax record data: A cautionary tale from Australia. *Journal of Economic Inequality*, 13(2):181–205.
- Burkhauser, R. V., Feng, S., Jenkins, S. P., and Larrimore, J. (2012). Recent trends in top income shares in the united states: Reconciling estimates from March CPS and IRS tax return data. *Review of Economics and Statistics*, 94(2):371–388.
- Burkhauser, R. V., Héroult, N., Jenkins, S. P., and Wilkins, R. (2016). 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.
- Chancel, L., Cogneau, D., Gethin, A., and Myczkowski, A. (2019). How Large Are African Inequalities? Towards Distributional National Accounts in Africa, 1990-2017. WID.World Working Paper 2019/13, World Inequality Lab.
- Chancel, L. and Piketty, T. (2019). Indian Income Inequality, 1922-2015: From British Raj to Billionaire Raj? *Review of Income and Wealth*, 65:S33–S62.
- Charpentier, A. and Flachaire, E. (2019). Pareto Models for Top Incomes. Technical report.
- Clementi, F. and Gallegati, M. (2005). Pareto’s law of income distribution: Evidence for Germany, the United Kingdom, and the United States. In Chatterjee, A., Yarlagadda, S., and Chakrabarti, B. K., editors, *Econophysics of Wealth Distributions*, chapter 1, pages 3–14. Springer, Milano.
- Cord, L., Genoni, M. E., and Rodríguez-Castelán, C. (2015). Overview. In Cord, L., Genoni, M. E., and Rodríguez-Castelán, C., editors, *Shared prosperity and Poverty Eradication in Latin America and the Caribbean*, chapter 1, pages 1–51. World Bank, Washington, DC.
- Cowell, F. A. and Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics*, 141(2):1044–1072.
- Davidson, R. and Flachaire, E. (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics*, 141(1):141–166.
- Deaton, A. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and statistics*, 87(1):1–19.
- Díaz-Bazán, T. (2015). Measuring Inequality from Top to Bottom. World Bank Policy Research Paper 7237, World Bank.
- Eckerstorfer, P., Halak, J., Kapeller, J., Schütz, B., Springholz, F., and Wildauer, R. (2016). Correcting for the missing rich: An application to wealth survey data. *Review of Income and Wealth*, 62(4):605–627.
- Ferreira, F. H. G., Messina, J., Rigolini, J., López-Calva, L. F., Lugo, M. A., and Vakis, R. (2013). *Economic mobility and the rise of the Latin American middle class*. The World Bank, Washington D.C.
- Flores, I., Sanhueza, C., Atria, J., and Mayer, R. (2019). Top incomes in Chile: A historical perspective on income inequality, 1964-2017. *Review of Income and Wealth*, Forthcoming.
- Gabaix, X. and Ibragimov, R. (2011). Rank- $1/2$: a simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, 29(1):24–39.

- Gasparini, L. (2004). Different lives: Inequality in Latin America. In De Ferranti, D., Perry, G. E., Ferreira, F. H., and Walton, M., editors, *Inequality in Latin America and the Caribbean: Breaking with History?*, chapter 2, pages 35–71. World Bank, Washington.
- Gasparini, L., Cruces, G., Tornarolli, L., and Mejía, D. (2011). Recent trends in income inequality in Latin America. *Economía*, 11(2):147–201.
- Genoni, M. E. and Salazar, M. (2015). Steering towards shared prosperity in Peru. In Cord, L., Genoni, M. E., and Rodríguez-Castelán, C., editors, *Shared Prosperity and Poverty Eradication in Latin America and the Caribbean*, chapter 8, pages 269–301. World Bank, Washington, DC.
- Hajargasht, G. and Griffiths, W. E. (2013). Pareto-lognormal distributions: Inequality, poverty, and estimation from grouped income data. *Economic Modelling*, 33:593–604.
- Herrera, J. (2017). Poverty and economic inequalities in Peru during the boom in growth: 2004-14. In Carbonnier, G., H. C. y. S. T. V., editor, *Alternative Pathways to Sustainable Development: Lessons from Latin America*, chapter 8, pages 138–173. Brill Nijhoff.
- Higgins, S., Lustig, N., and Vigorito, A. (2018). The rich underreport their income: Assessing bias in inequality estimates and correction methods using linked survey and tax data. ECINEQ WP 2018-475, ECINEQ.
- Ibragimov, M. and Ibragimov, R. (2018). Heavy tails and upper-tail inequality: The case of Russia. *Empirical Economics*, 54(2):823–837.
- Jenkins, S. P. (2017). Pareto models, top incomes and recent trends in UK income inequality. *Economica*, 84(334):261–289.
- Kim, N.-N. (2018). Top Incomes in Korea: Update, 1933-2016. *Korean Economic Forum*, 11(1):1–21.
- Lakner, C. and Milanovic, B. (2016). Global income distribution: from the fall of the Berlin Wall to the Great Recession. *World Bank Economic Review*, 30(2):203–232.
- Lustig, N. (2019). The ‘missing rich’ in household surveys: Causes and correction approaches. Working Paper 75, Commitment to Equity (CEQ) Institute, Tulane University.
- Lustig, N., López-Calva, L. F., and Ortiz-Juarez, E. (2013). Declining inequality in Latin America in the 2000s: The cases of Argentina, Brazil, and Mexico. *World Development*, 44:129–141.
- Morgan, M. (2018). 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.
- Novokmet, F., Piketty, T., Yang, L., and Zucman, G. (2018). From Communism to Capitalism: Private versus Public Property and Inequality in China and Russia. In *AEA Papers and Proceedings*, volume 108, pages 109–13.
- Paz, P. and Urrutia, C. (2015). Economic growth and wage stagnation in Peru: 1998-2012. *Review of Development Economics*, 19(2):328–345.
- Piketty, T. (2003). Income inequality in France, 1901-1998. *Journal of Political Economy*, 111(5):1004–1042.
- Piketty, T. and Saez, E. (2003). Income inequality in the United States, 1913-1998. *Quarterly Journal of Economics*, 118(1):1–41.
- Piketty, T., Saez, E., and Zucman, G. (2018). Distributional national accounts: Methods and estimates for the United States. *Quarterly Journal of Economics*, 133(2):553–609.

- Piketty, T., Yang, L., and Zucman, G. (2019). Capital accumulation, private property, and rising inequality in China, 1978-2015. *American Economic Review*, 109(7):2469–96.
- Reed, W. J. (2003). The Pareto law of incomes: An explanation and an extension. *Physica A: Statistical Mechanics and its Applications*, 319:469–486.
- Reed, W. J. and Jorgensen, M. (2004). The double Pareto-lognormal distribution—a new parametric model for size distributions. *Communications in Statistics - Theory and Methods*, 33(8):1733–1753.
- Ruiz, N. and Woloszko, N. (2016). What do household surveys suggest about the top 1% incomes and inequality in OECD countries? OECD Economics Department Working Paper 1265, OECD.
- Stampini, M., Robles, M., Sáenz, M., Ibarrarán, P., and Medellín, N. (2016). Poverty, vulnerability, and the middle class in Latin America. *Latin American Economic Review*, 25(1):Article 4.
- Székely, M. and Mendoza, P. (2016). Declining inequality in Latin America: Structural shift or temporary phenomenon? *Oxford Development Studies*, 45(2):204–221.
- Vermeulen, P. (2018). How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, 64(2):357–387.
- Winkelried, D. and Torres, J. (2019). Economic mobility along the business cycle. the case of Peru. *Applied Economics*, 51(18):1894–1906.
- Yamada, G., Castro, J. F., and Bacigalupo, J. (2012). Desigualdad monetaria en un contexto de rápido crecimiento económico: El caso reciente del Perú. *Revista Estudios Económicos*, 24:65–77.

Figures

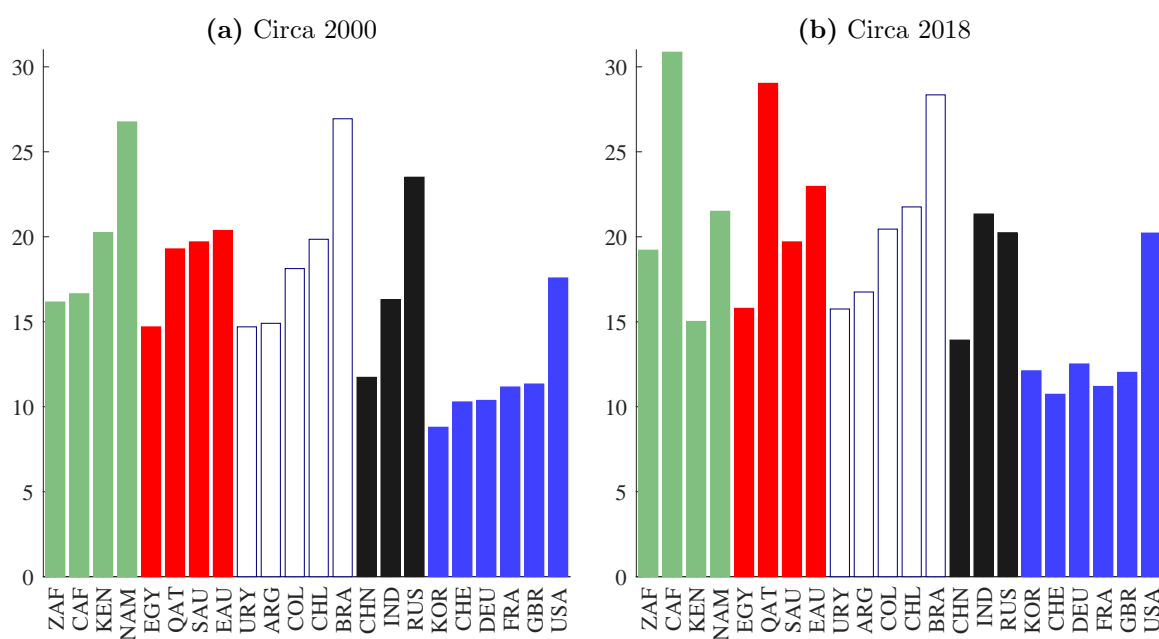
Figure 1. *Gini indices in the world and in Latin America, 2000 to 2018*



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).

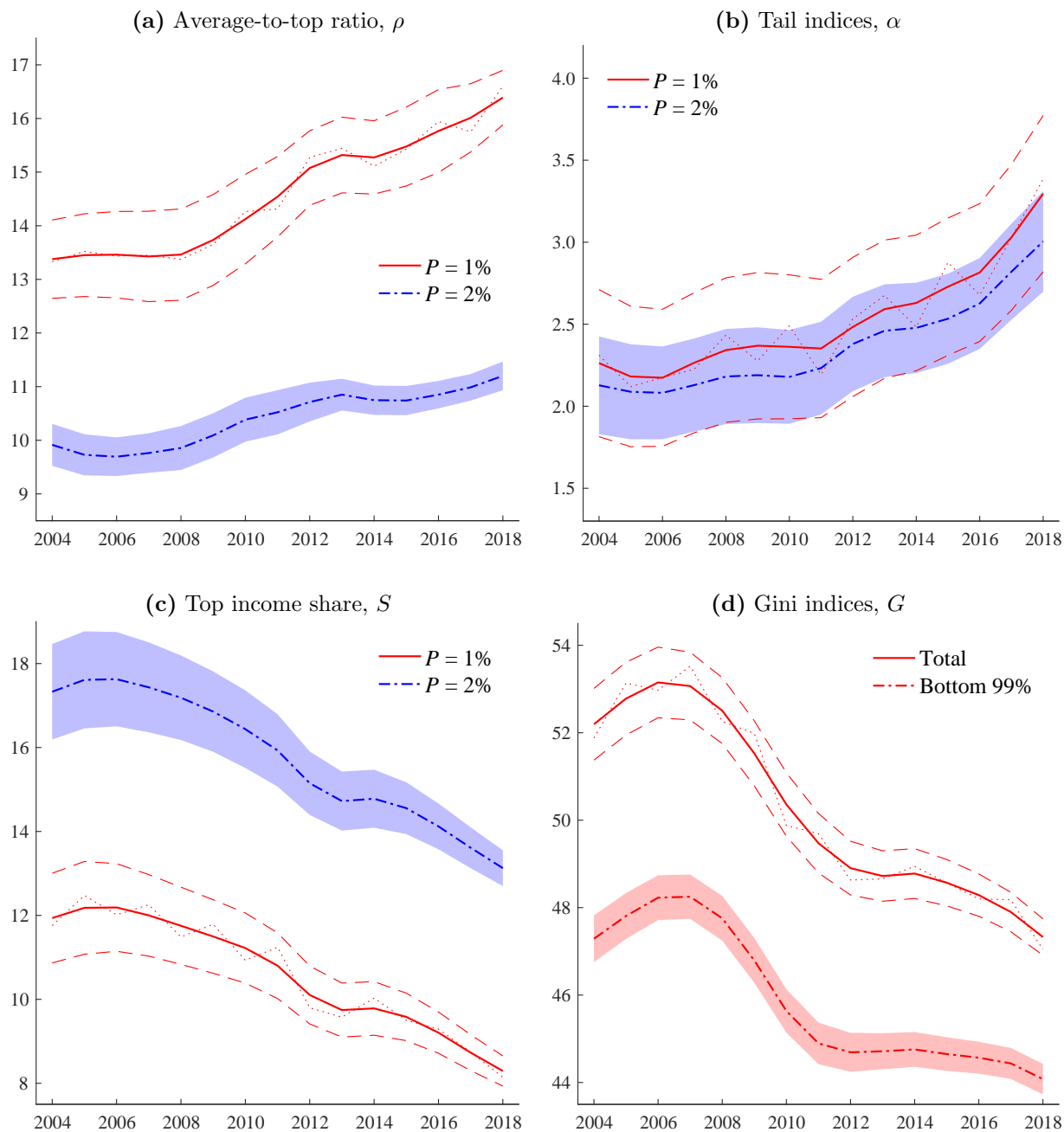
Figure 2. Top 1% income shares for selected countries, 2000 to 2018



Source: WID.World. 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.

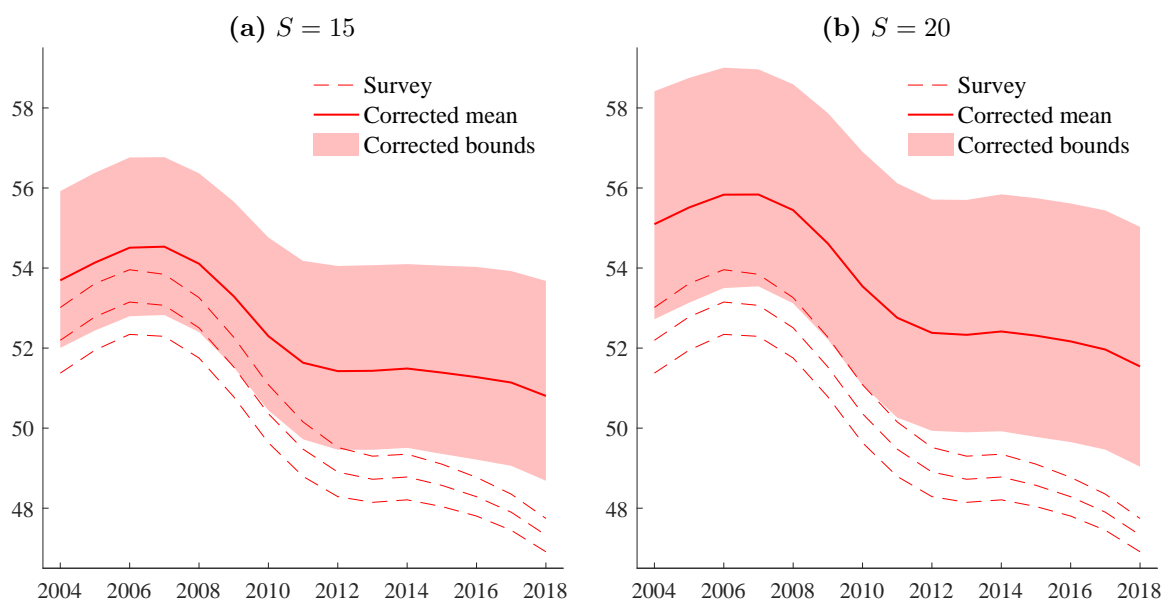
Figure 3. *Income inequality in the Peruvian household surveys, 2004 to 2018*



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.

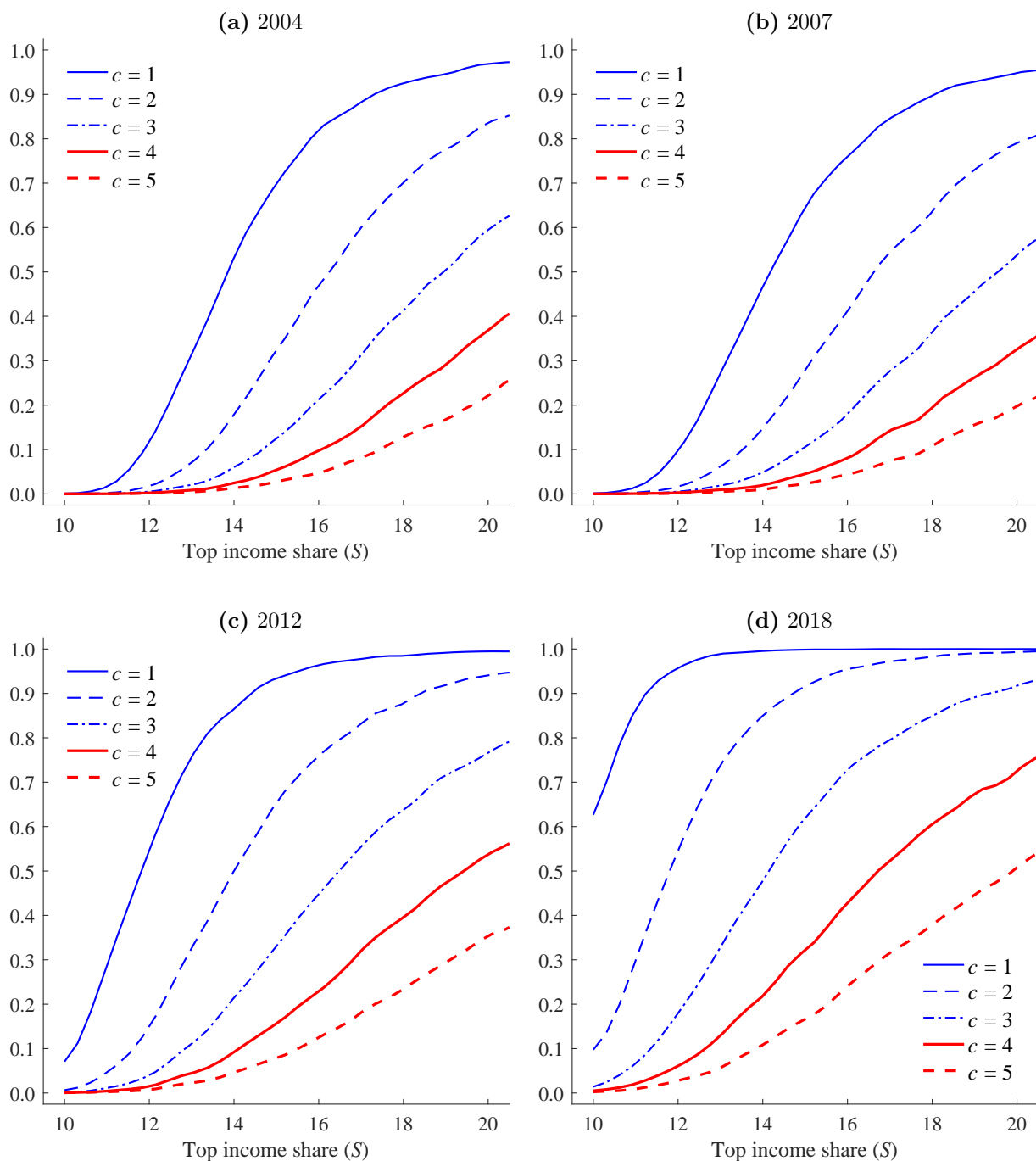
Figure 4. *Sample and corrected Gini indices, 2004 to 2018*



Source: ENAHO, rounds from 2004 to 2018. Own elaboration.

Notes: Sample and simulated Gini indices. The shaded areas represent a 90% interval of the corrected Gini indices and the continuous line is the mean across 5,000 replications.

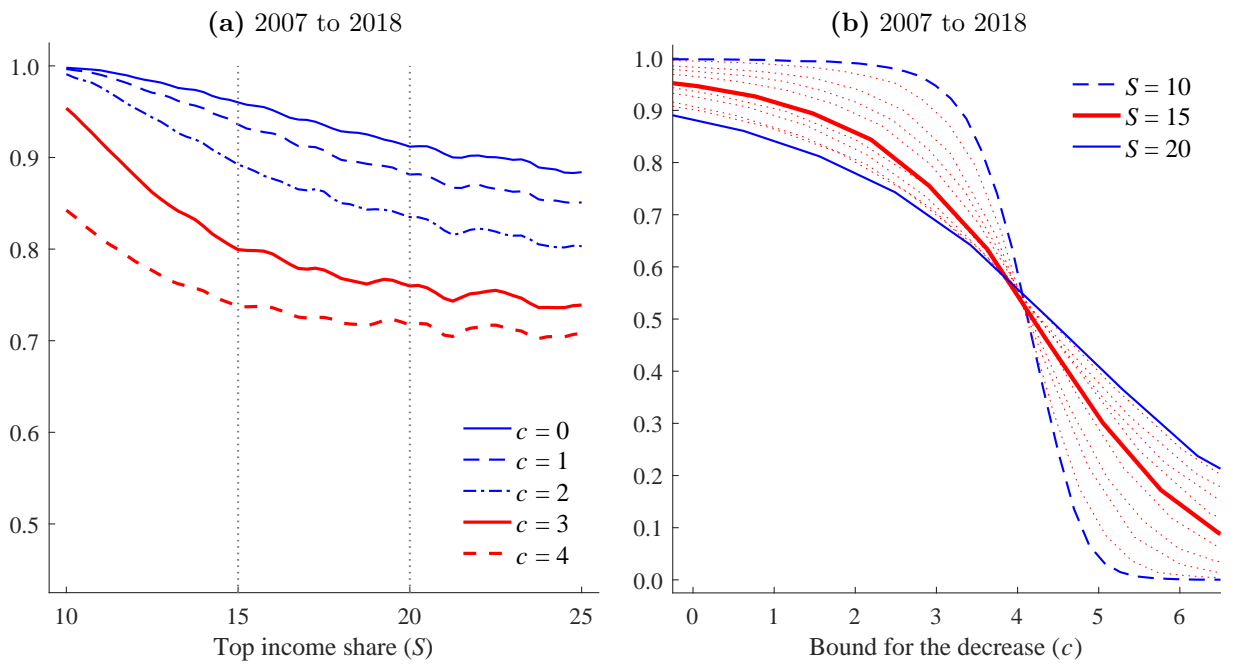
Figure 5. Probabilities of correction in the Gini index by year, 2004 to 2018



Source: ENAHO, rounds from 2004 to 2018. Own elaboration.

Notes: The graphs show $\Pr(G - \bar{G} > c)$ 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.

Figure 6. Probabilities of change in the Gini index, 2007 to 2018

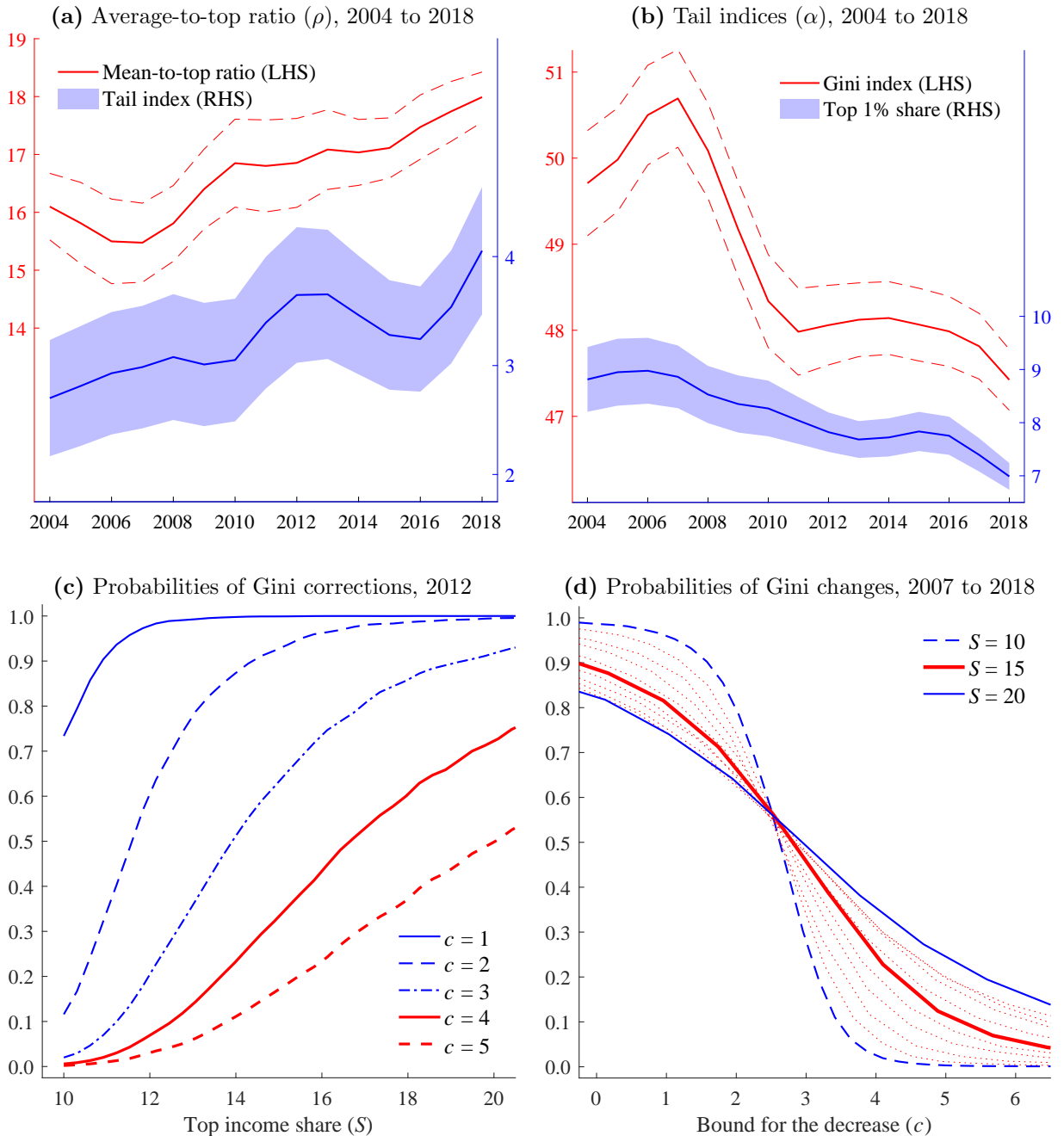


Source: ENAHO, rounds from 2004 to 2018. Own elaboration.

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.

Online supplement (not intended for publication):
Results for alternative measures of income

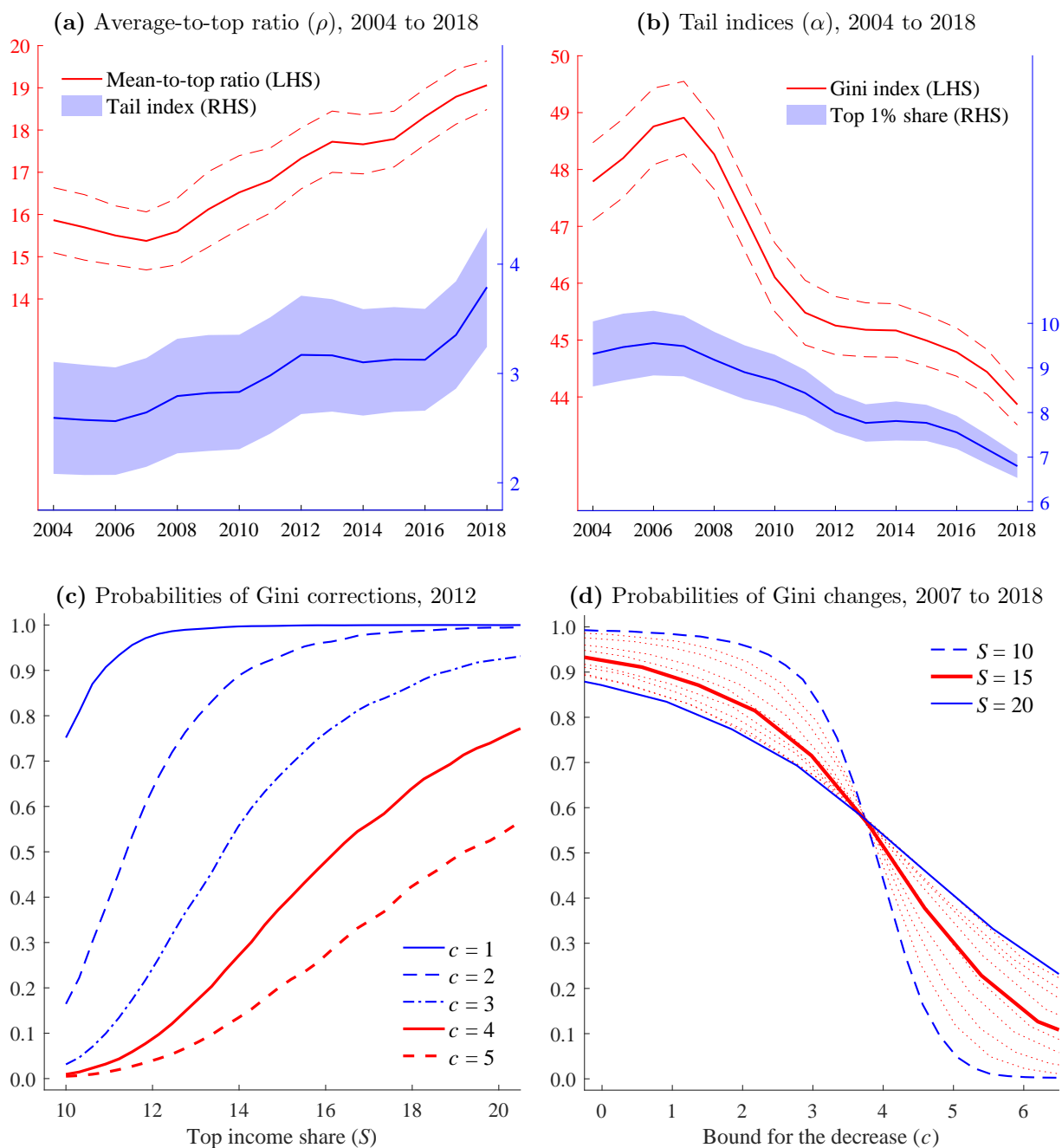
Figure A1. Results for total household income before taxes



Source: ENAHO, rounds from 2004 to 2018. Own elaboration.

Notes: Total household income before taxes is reported directly in ENAHO with the code `inghog1d`. For panels (a) and (b), see the notes to Figure 3 in the main text; for panel (c), the notes to Figure 5; for panel (d), the notes to Figure 6.

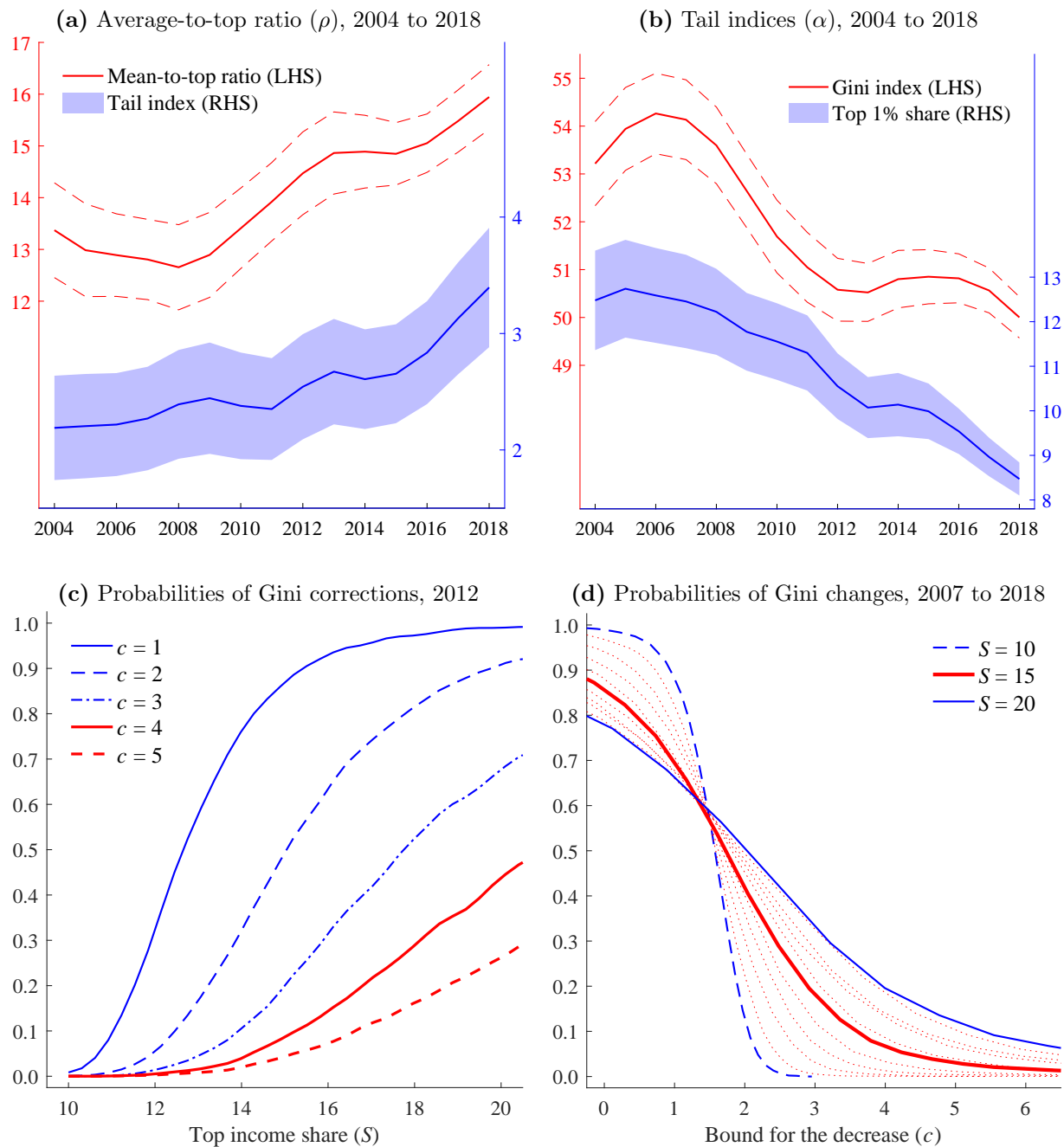
Figure A2. Results for (square root) scale equivalent income before taxes



Source: ENAHO, rounds from 2004 to 2018. Own elaboration.

Notes: Scale equivalent income before taxes is the variable coded `inghog1d` (total household income before taxes) divided by the square root of the variable coded `mieperho` (total number of household members). For panels (a) and (b), see the notes to Figure 3 in the main text; for panel (c), the notes to Figure 5; for panel (d), the notes to Figure 6.

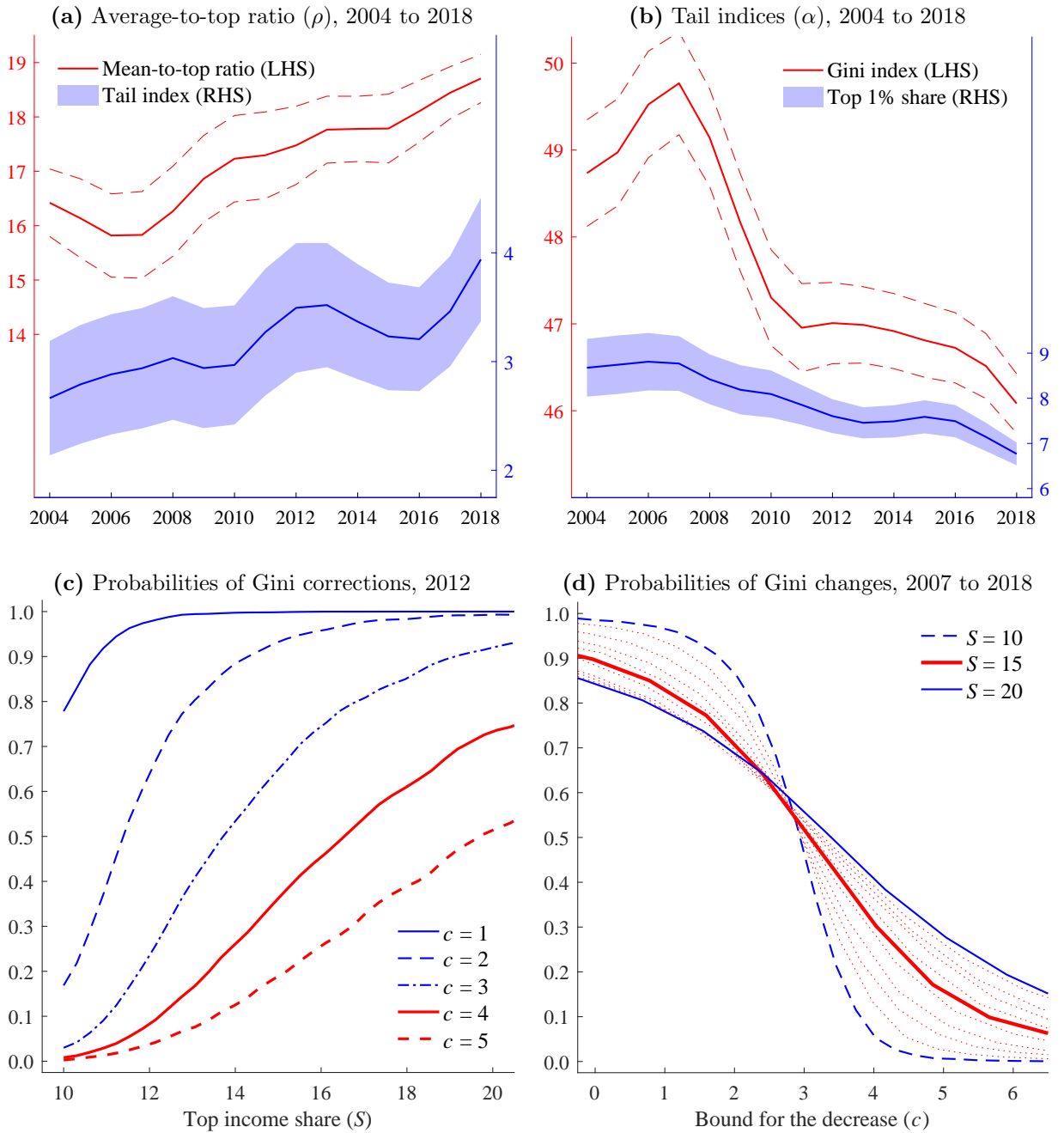
Figure A3. Results for per capita labor income before taxes



Source: ENAHO, rounds from 2004 to 2018. Own elaboration.

Notes: Per capita labor income before taxes is the sum of before-tax incomes earned from dependent work or self-employment, as the main or as a secondary job, for all employed household members (module 500 of ENAHO on wages and employment), and divided by the total number of household members (coded `mieperho`). For panels (a) and (b), see the notes to Figure 3 in the main text; for panel (c), the notes to Figure 5; for panel (d), the notes to Figure 6.

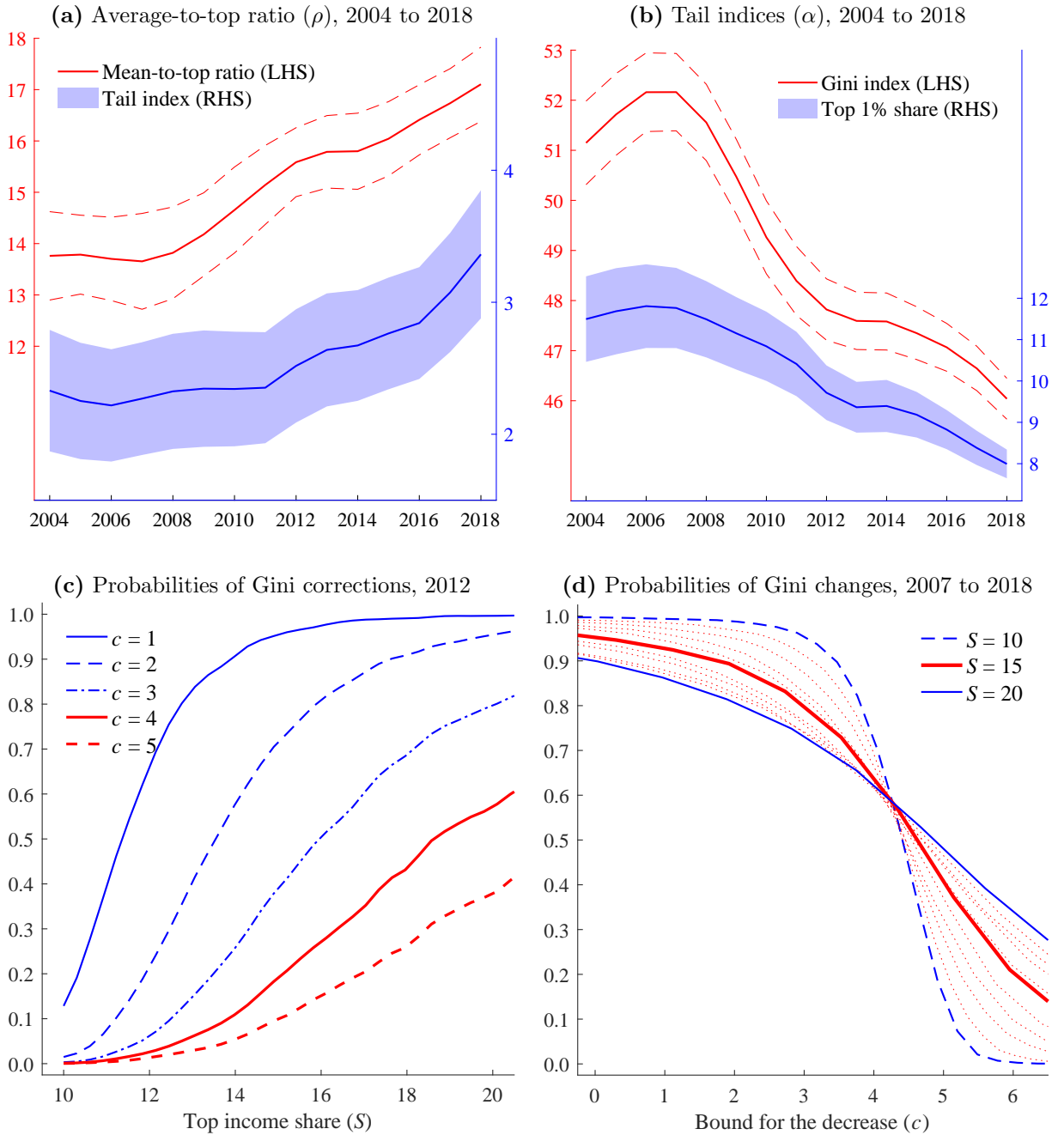
Figure A4. Results for total household income after taxes



Source: ENAHO, rounds from 2004 to 2018. Own elaboration.

Notes: Total household income after taxes is reported directly in ENAHO with the code `inghog2d`. For panels (a) and (b), see the notes to Figure 3 in the main text; for panel (c), the notes to Figure 5; for panel (d), the notes to Figure 6.

Figure A5. Results for per capita income after taxes

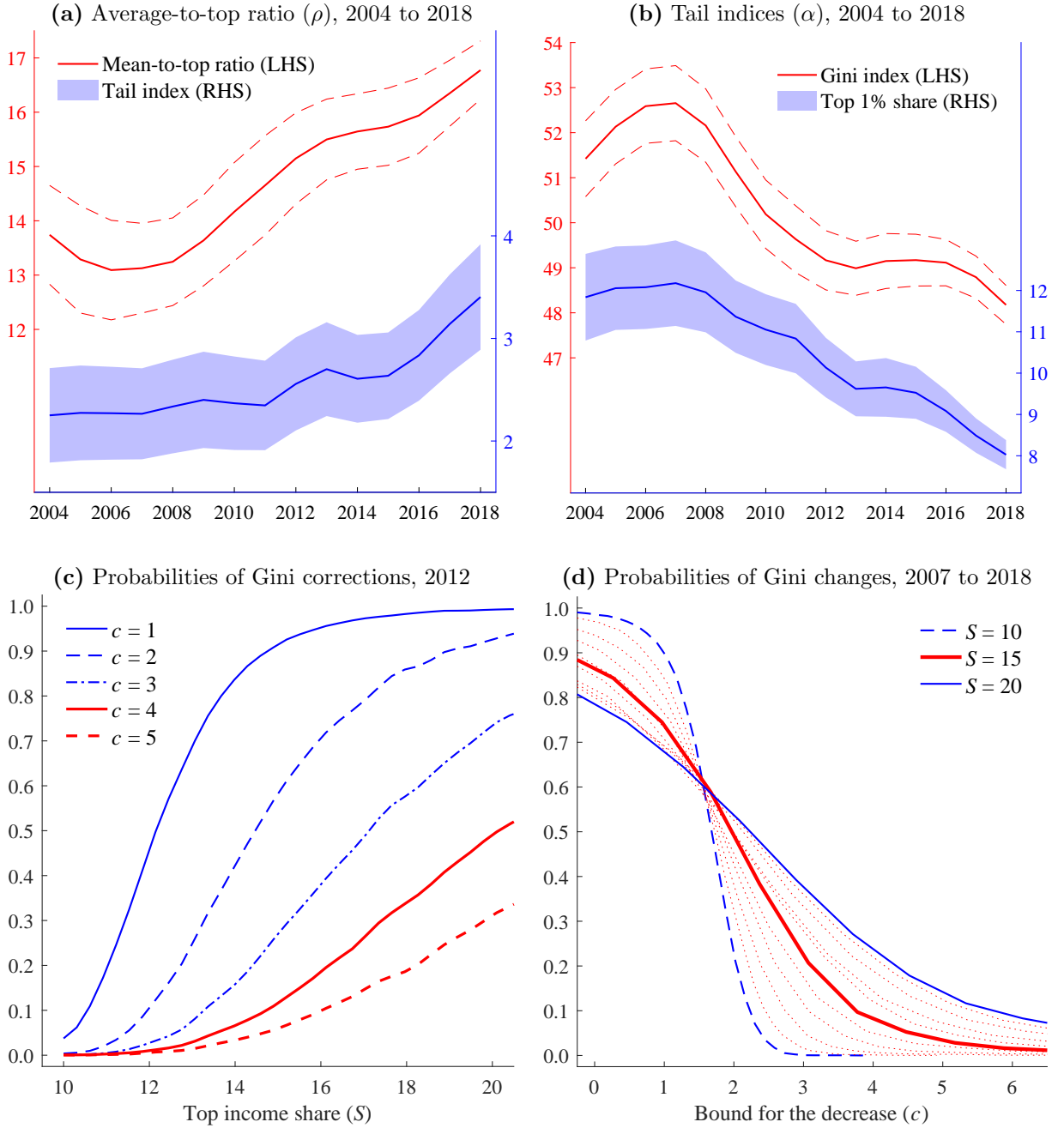


Source: ENAHO, rounds from 2004 to 2018. Own elaboration.

Notes: Per capita household income after taxes is the variable coded `inghog2d` (total household income after taxes) divided by the variable coded `mieperho` (total number of household members).

For panels (a) and (b), see the notes to Figure 3 in the main text; for panel (c), the notes to Figure 5; for panel (d), the notes to Figure 6.

Figure A6. Results for per capita labor income after taxes



Source: ENAHO, rounds from 2004 to 2018. Own elaboration.

Notes: Per capita labor income after taxes is the sum of after-tax incomes earned from dependent work or self-employment, as the main or as a secondary job, for all employed household members (module 500 of ENAHO on wages and employment), and divided by the total number of household members (coded `mieperho`). For panels (a) and (b), see the notes to Figure 3 in the main text; for panel (c), the notes to Figure 5; for panel (d), the notes to Figure 6.