the review of income and wealth

Review of Income and Wealth
Series 69, Number 4, December 2023
DOI: 10.1111/roiw.12618

UNDERREPORTING OF TOP INCOMES AND INEQUALITY: A COMPARISON OF CORRECTION METHODS USING SIMULATIONS AND LINKED SURVEY AND TAX DATA

by Emmanuel Flachaire

Aix-Marseille Université, AMSE

Nora Lustig

Tulane University

and

Andrea Vigorito

Instituto de Economía, FCEA, Universidad de la República (Uruguay)

Household surveys do not capture incomes at the top of the distribution well. This yields biased inequality measures. We compare the performance of the reweighting and replacing methods to address top incomes underreporting in surveys using information from tax records. The biggest challenge is that the true threshold above which underreporting occurs is unknown. Relying on simulation, we construct a hypothetical true distribution and a "distorted" distribution that mimics an underreporting pattern found in a novel linked data for Uruguay. Our simulations show that if one chooses a threshold that is not close to the true one, corrected inequality measures may be significantly biased. Interestingly, the bias using the replacing method is less sensitive to the choice of threshold. We approach the threshold selection challenge in practice using the Uruguayan linked data. Our findings are analogous to the simulation exercise. These results, however, should not be considered a general assessment of the two methods.

JEL Codes: C81, D31

Keywords: correction methods, household surveys, income underreporting, inequality, linked data, replacing, reweighting, tax records

Note: We thank Uruguay's Instituto Nacional de Estadística and Dirección General Impositiva for providing the data used in this research. We are very grateful to Sean Higgins for his contributions at earlier stages of this project. Two anonymous peer reviewers provided very useful comments, we thank them for that. We are also grateful for very useful comments from François Bourguignon and other participants of the workshop "Methodological Advances in Fiscal Incidence Analysis: Commitment to Equity Institute" (Universidad de San Andres, Buenos Aires, 2017); participants in LACEA 2017 annual conference; participants of the "Workshop on harmonization of household surveys, fiscal data and national accounts: comparing approaches and establishing standards" (Paris School of Economics, 2018); participants of Instituto de Economía Seminar (2018); participants of ECINEQ 2019; Thomas Blanchet, Mauricio De Rosa, Ignacio Flores, Johannes Koenig, Marc Morgan, Christian Schluter, and Joan Vilá; and Siyu Quan for research assistance. Emmanuel Flachaire thanks the financial support from the projects ANR-17-EURE-0020, ANR-17-CE41-0007, and ANR-19-FRAL-0006 managed by the French National Research Agency and from the Excellence Initiative of Aix-Marseille University–A*MIDEX. Nora Lustig acknowledges the generous support of the Bill & Melinda Gates Foundation (grant number: OPP1135502). Andrea Vigorito received financial support from Universidad de la República (Uruguay).

*Correspondence to: Andrea Vigorito, Instituto de Economía, FCEA, Universidad de la República (Uruguay), Av. Gonzalo Ramírez 1926, 11200 Montevideo, Uruguay (andrea.vigorito@fcea.edu.uy).

1. Introduction

Household surveys suffer from representation errors, errors due to item and unit nonresponse, and measurement errors.[1] Such errors can affect the entire survey, but here we are particularly concerned when they occur in the upper tail. Household surveys do not capture incomes at the top of the distribution well because the rich may be harder to reach, leading to unit nonresponse; more likely to refuse to answer when reached, resulting in item nonresponse; or may report a lower fraction of their income when responding to the survey, resulting in underreporting (Atkinson, 2007). In addition, in finite samples the upper tail is not captured well due to sparseness or because data producers truncate or top code the distributions in the upper tail (Cowell and Flachaire, 2007, 2015; Biemer and Christ, 2008). These issues can lead to significant bias in inequality measures, and this bias can be either positive or negative (Deaton, 2005). Recognizing this, throughout the years researchers have resorted to using other sources of information to correct survey data or survey-based inequality estimates. These other sources include National Accounts (Altimir, 1987; Piketty *et al.*, 2018), administrative data from tax and social security records (Burkhauser *et al.*, 2016; Jenkins, 2017; Piketty *et al.*, 2019), complementary surveys (Fisher *et al.*, 2022), and the so-called rich lists (Brzezinski, 2014). For a survey, see Lustig (2019).

Here we focus on comparing correction methods to address one type of measurement error: underreporting of income in the upper tail.[2] Two main approaches have been used in the literature to correct survey upper-tail errors, including underreporting. Following Hlasny and Verme (Hlasny and Verme, 2018, pp. 1–2), we call these approaches “replacing” and “reweighting” (see details on these two methods in Section 3). Both correction approaches rely on implicit assumptions that are often untestable.[3] In particular, they rely on the appropriate selection of the threshold beyond which survey data tend to underreport income. The biggest challenge in applying correction methods is that the true income distribution is unknown; therefore, one does not know the threshold above which underreporting occurs.

To analyze the sensitivity of correction methods to the choice of threshold, we rely on simulation. The approach allows us to focus on underreporting and not consider sampling errors in the upper tail, a common problem in finite samples. We simulate a hypothetical true distribution and a “distorted” distribution that suffers from underreporting. (Relying on hypothetical distributions has the additional advantage that we can focus on underreporting and not consider sampling errors in the upper tail, a common problem in finite samples). The distorted distribution is not just arbitrarily constructed. It mimics an actual pattern of underreporting

[1] The total survey error is composed of the sum of three distinct elements: representation error, error due to non-response, and measurement error (Groves and Lyberg, 2010; Meyer and Mittag, 2019).

[2] We do not address the case in which the entire income is not reported (item nonresponse). Because most likely the behavior underlying item nonresponse is different from misreporting, this would warrant a separate type of analysis.

[3] In fact, the survey earnings validation literature concludes that the definition of a true distribution largely depends on priors chosen by researchers, which lead to different measurement error estimates (Kapteyn and Ypma, 2007; Abowd and Stinson, 2013; Jenkins and Rios-Avila, 2020). See also Gottschalk and Huynh (2010), Hyslop and Townsend (2020), and Adriaans *et al.* (2020).

found in a novel data set that directly links a subset of individuals from Uruguay's official household survey to the same individuals' tax returns, enabling us to observe income reported from each of these sources for the same person.[4] We find that underreporting in the survey occurs primarily in the upper part of the tax records income distribution and underreporting increases with income. The latter is also found in previous survey earnings validation studies for developed countries such as Abowd and Stinson (2013) and Adriaans *et al.* (2020).

These studies raise some hypotheses for why individuals underreport their income, such as prevailing social norms, informality, cognitive problems, and fear of penalties by the tax authorities. It could also be due to the use of proxy respondents. With the available data, however, we are not able to disentangle the reasons for this behavior. The next step is to calculate the Gini coefficient, the mean log deviation (MLD), the Theil index, and the top 10 percent, 5 percent, and 1 percent shares for both the true and distorted distributions and find that all the inequality indicators estimated for the latter are strongly biased. We then apply the two correction methods—replacing and reweighting—for a series of thresholds, including those used in the literature.

Our analysis shows that threshold selection plays a key role (see Cowell and Flachaire, 2015 for a discussion of the challenges around threshold selection). If the threshold is not close to the true threshold, inequality measures may be significantly biased. An interesting finding is that, in the case of underreporting, the replacing method is less sensitive to the choice of the threshold.[5] In other words, the replacing method yields inequality measures that are closer to the true inequality measures for a broader set of thresholds than reweighting. This is because with replacing, the error introduced with corrections is confined to a smaller segment of the distribution. In contrast, reweighting affects the entire distribution below the threshold and thus—unless one chose the correct threshold—–reweighting may introduce biases into inequality measures that are sensitive to the bottom of the distribution. If one knows the correct threshold, replacing and reweighting are equivalent and applying either would yield the true distribution. However, the challenge is precisely that, in practice, the true threshold remains unknown.

In addition to the simulations, we explore how to approach the threshold selection challenge in practice using the linked data for Uruguay. Our results are analogous to our simulation exercise: with replacing, the inequality measures are less sensitive to threshold selection. These results, however, should not be considered a general assessment of the two approaches. "Reweighting" and "replacing" are two

[4]The linked data are restricted to adults in households with children aged 0–3. Although this subsample captures households at the top of the income distribution, this is probably a biased sample if we were interested in measuring the distribution of income in Uruguay. However, the purpose of using this linked data is not to estimate inequality in Uruguay, at least not here. We use the linked data to observe an actual pattern of underreporting; such an observation is usually not possible as there are very few instances for which linked data exist.

[5]It is worth pointing out that observing that inequality measures have converged to a stable value cannot be interpreted as suggesting that one has found the threshold that is closest to the correct one. That is, while convergence is a necessary condition to approximating the correct threshold, it is not a sufficient one.

broad classes of methods, among which some specific applications of replacing and reweighting may outperform others. Therefore, one particular replacing method may outperform one particular reweighting method, but this ranking may not hold for all replacing/reweighting methods.

This article is organized as follows. We first describe the databases used in this study and show the misreporting patterns identified in the linked data (Section 2.2). We then present the correction methods and provide simulation results (Section 3). Based on these findings, Section 4 discusses what to do in practice, and to illustrate, presents corrected inequality measures estimated with the linked data for Uruguay. Section 5 includes some final remarks. Additional information can be found in the Appendix.

2. Misreporting Evidence from Linked Data

2.1. *Data*

We use a novel database in which a subsample of Uruguay's official household survey—*Encuesta Continua de Hogares* (ECH)—has been linked to personal income tax records from the *Dirección General Impositiva* (DGI) by Instituto Nacional de Estadistica preserving statistical secrecy and confidentiality.

ECH collects post-tax information on labor income and social security coverage (formality) for each worker, separately considering (1) self-employment earnings, (2) main salaried occupation, and (3) remaining salaried occupations. Based on this information, we compute post-tax formal labor income by adding all the taxable income components in the survey: post-tax salaries and wages, post-tax commissions, incentives, vacation pay, and overtime payments. We then add post-tax pensions and post-tax capital income to obtain total post-tax income. That is, the survey post-tax income is equivalent to post-tax income from tax records. In our analysis, we exclude tips, arrears, transport, food or housing vouchers, other in-kind payments, other fringe benefits, and bonuses from formal occupations, but income misreporting patterns remain unchanged if these income categories are considered (Higgins *et al.*, 2018).

DGI databases used in this study include the universe of potential personal income tax payers, including all formal workers, pensioners, self-employed (liberal) professionals, and capital income recipients for 2012–2013 (Burdín *et al.*, 2014). As a whole, these data cover approximately 75 percent of the population aged 20 or more. Like all tax records, these data are subject to tax evasion and avoidance (Atkinson, 2007).[6]Personal income taxation (*Impuesto a las Retribuciones de las Personas Físicas*, IRPF) in Uruguay is based on a dual scheme that combines a progressive tax schedule for labor income and pensions, with a flat tax rate on capital income. The tax unit is the individual, but married couples have the option of filing a joint tax return for labor income. However, only 1.8 percent of individuals in the tax records chose this regime and, in fact, we do not have couples under this

[6]In the 2000s, Uruguay experienced rapid economic growth, coupled with a substantial decrease in informal employment: from 40 percent in 2004 to 23.5 percent in 2014 (Carrasco *et al.*, 2018). Thus, formal workers represent the majority of total workers.

tax regime in our merged sample. Personal income tax is withheld, reported, and paid by employers, firms, banks, and other agents. Only individuals with more than one occupation or those who receive more than one income source, as well as the self-employed, file taxes (these individuals make up 5 percent of the observations included in our merged sample). More information on the tax scheme can be found in Appendix 1.

A subsample of individuals included in the 2012–2013 ECH was linked to their tax records. The subsample of linked individuals consists of those who were included in a follow-up survey: the Nutrition, Child Development, and Health Survey (*Encuesta de Nutrición, Desarrollo Infantil, y Salud*, ENDIS). ENDIS is a longitudinal study that follows 2,649 urban households with children aged 0–3 who were interviewed as part of ECH between February 2012 and August 2013 (Instituto Nacional de Estadística, 2013, 2018). The potentially linked individuals include all adult members (aged 18 years or more) of a household. By the time of the writing of this article, there had been three waves of ENDIS. In the first wave (started in November 2013), enumerators collected the unique national identification number (*cedula*) of each respondent (principal caregivers of reference children, mainly mothers), whereas in the second wave (2015), this information was also gathered for fathers and other adult household members, allowing INE and DGI to merge all adults (not just mothers) from the 2012–2013 ECH who were in ENDIS to DGI tax records. *Cedulas* are composed of seven digits and a verification number. As INE gathered the verification number, it ensured that the numbers provided were correct, minimizing potential linkage errors. We did not have access to the actual card numbers, only to masked identifiers.

With the linked data, we can compare tax-return incomes with survey incomes for the same individuals. To compare incomes reported in ECH with DGI, for each individual we create harmonized post-tax total income variables with ECH and DGI data, by adding formal labor income, pensions, and taxable capital income (rents, dividends, entrepreneurial profits, and interest from bank deposits and other financial assets). In ECH, the reference period for the collection of data on labor earnings and pensions is the previous month; in DGI, we consider information corresponding to the month before the ECH interview. Capital income is collected in the two databases on an annual basis, and thus we include a monthly average.

From the original ENDIS households, 4,539 adults were income receivers and 2,360 were formal workers (whose incomes are positive or zero in the period of reference) or received income from capital or pensions. Of these, 2,287 had valid ID cards (either the interviewee or other adults in the household) and were linked to tax data. Among linked observations, 1,634 (71 percent) had positive earnings in tax records in the month before the ECH interview. Of those 1,634, a total of 1,471 had positive income in ECH, and a considerably lower proportion (163, 10 percent) did not report their income but had positive income in DGI records.[7] In line

[7]Among the 2,749 survey respondents, 1,720 declared labor force participation (1,507 employed and 213 unemployed). The remainder were housekeepers/homemakers (808), full-time students (194), pensioners, or rentiers. A total of 1,079 were formal workers, rentiers, or received pensions, and 1,027 were merged to the DGI database (95 percent). In addition, 265 individuals report zero income in tax records but not in the survey.

TABLE 1
Descriptive Statistics

Statistic	ECH Income	DGI Income	*m* (ECH-DGI)
Mean	19,363	22,585	−3,221
SD	164,560	34,260	28,314
Min.	1,062	82	−919,153
Max.	191,185	979,153	143,488
Correlation coefficient			
ECH	1		
DGI* (*)	0.569	1	
m (ECH−DGI)	−0.106	−0.878	1

(*) Considering labor income only, the correlation coefficient rises to 0.667.

Notes: Harmonized monthly post-tax income. Linked cases. Descriptive statistics for harmonized post tax income (in Uruguayan currency) were computed with ECH and DGI data for the subset of linked observations that had harmonized income different from zero in the two data sets in the survey reference period. 1 US Dollar = 20.45 Uruguayan pesos.

Source: Authors' calculations based on ECH, ENDIS, and DGI microdata.

with the earnings validation literature (Bollinger *et al.*, 2019), the latter figure can be considered a measure of item non-response in the survey. Figure A.2.1 depicts the proportion of item non-response among individuals by percentile of DGI tax records, which heavily accumulates in the lower tail of the income distribution, clearly rejecting the missing at random hypothesis. This pattern is different from the two tails pattern identified by Bollinger *et al.* (2019) for the US. As our exercise focuses on misreporting, we use the 1,471 linked individuals who reported positive income in both sources. Among linked observations, average income is 17 percent larger in tax records than that in survey reports and, perhaps unsurprisingly, maximum income is higher in the tax data (Table 1). Restricting the comparison to labor income yields similar results for average income, although the DGI maximum is 25 percent higher than the ECH maximum.

Our analysis is restricted to couples with children aged 0–3 living in urban areas. As shown in Figure 1, this subsample of linked observations is present along the whole ECH and the full DGI income distributions and the proportion of observations is even larger in the upper strata. This means that our subsample can be used as an adequate approximation of the patterns of misreporting that could potentially be observed in the full ECH. For a comparison of the characteristics of linked observations and the rest of the individuals in ENDIS, see Table A.2.1.

2.2. *Misreporting Patterns*

To analyze the measurement error in our linked data, we examine a subsample of 1,471 individuals who have positive income in both the ECH and DGI during the period of reference. Figure 2 plots the ratio of income reported in tax data to income reported in the survey for each observation in the linked data, and shows how this varies across the tax return distribution (i.e. by tax return income percentile). A local linear regression is estimated with a bandwidth obtained by cross-validation and with bootstrap standard errors. It is clear from this figure that individuals tend to underreport their income in the survey, above the minimum taxable income. If everyone reported the same income in the two data sources, all points would lie

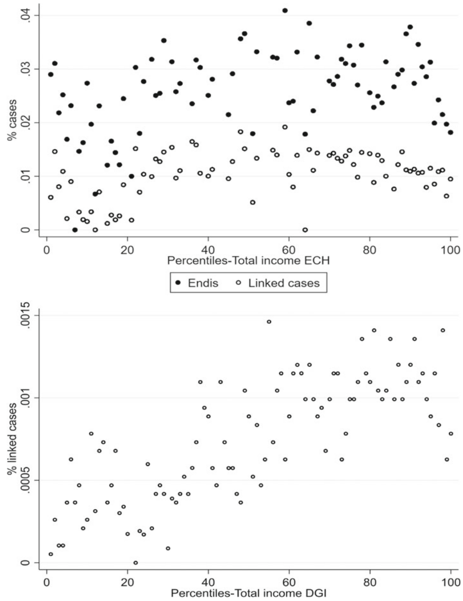

Figure 1. Proportion of ENDIS Income Receivers and Linked Observations by Income Percentile in ECH and DGI.

Notes: In the first panel, the label *ENDIS* corresponds to adults with positive income in ECH. Percentiles were built with the full set of ECH adults receiving positive income. In the second panel, percentiles were built with the full set of 2012/2013 DGI observations.

Source: Authors' calculations based on ECH and ENDIS microdata

Figure 2. Linked Data: Misreporting Rates.

Notes: Misreporting rates are computed as Ratios of Tax Return to Survey Income (Circles), with Nonparametric Estimation of Average Misreporting Rates (Red Line). A local linear regression is estimated with a bandwidth obtained by cross-validation and with bootstrap standard error with the `npreg` function in `R`.

Source: Authors' calculations based on ECH and ENDIS microdata. [Colour figure can be viewed at wileyonlinelibrary.com].

along the $y = 1$ line. If incomes were reported with noise but income misreporting was orthogonal to income, the points would bounce around, but the average relationship would correspond to $y = 1$. Figure A.2.2 shows the empirical copula (i.e. bivariate density) of percentiles in the survey and in tax return income distributions. We also fit a local quantile regression at the median, and we find a similar pattern (not reported), so underreporting is not driven by a few outliers.

It can be observed that ECH incomes exceed tax return incomes in approximately the bottom half of the tax return distribution, while survey incomes are lower in the top half. Interestingly, the point at which survey incomes are lower than tax return incomes corresponds to the minimum taxable income threshold for labor earnings, which represent the largest income component both in this subsample and in the full tax records distribution. In this case, the survey reporting pattern we obtain (overreporting in the lower tail and underreporting at the top) is in line with previous findings from the survey earnings validation literature (Adriaans *et al.*, 2020).[8] The proportion of observations below the minimum income threshold in the linked sample is similar to that obtained for the full tax records distribution.[9] In fact, the top 1 percent of the full tax return distribution reports only about 60 percent of the income from their tax returns in the household survey. It is worth noting what happens in the low percentiles, for which the misreporting ratio can take on very small values. The confidence bands suggest that the ratio is significantly smaller than 1 below the minimum wage, for which tax data are known to be unreliable (Atkinson, 2007; Burkhauser *et al.*, 2016; Piketty *et al.*, 2019). To some extent, we can question the reliability of tax data under the value of the minimum taxable income, because income from informal employment is more prevalent below the median. However, overreporting below the minimum taxable income has little consequence for the purposes of this article. To check whether the misreporting pattern we identify in Figure 2 holds for different population groups, we compute the misreporting ratios for different income variables and population groups. We first restrict merged cases to harmonized labor income only, leaving aside the remaining income sources. Second, we consider full-time workers only, assuming that their income is more stable, and that they are less likely to misreport. Third, we consider only survey respondents, assuming that they present a higher probability of providing accurate responses (although in our regression analysis, the proxy-respondent variable was not statistically significant). Finally, we consider total income reported by each merged observation in the survey, including informal income from different occupations. As can be seen in Figure A.2.4, the results present slight variations, but

[8]In addition to social norms (a factor identified in the literature), overreporting at the bottom in the context of Uruguay (as well as more generally in low- and middle-income countries) is likely the result of the coexistence of income coming from both formal and informal employment. It is worth pointing out that the proportion of income coming from informal occupations is between 0 percent and 15 percent among linked cases (Figure A.2.3), and high-income individuals also report informal income. To further investigate misreporting behavior, we regress the survey/tax income (ECG/DGI) ratio against a set of observable characteristics of the individual (Table A.2.3).

[9]Because in Latin American countries the tax system relies heavily on indirect taxation, the minimum taxable labor income threshold in Uruguay is set at a high quantile of the taxable income distribution compared to European countries. Existing studies rule out bunching and other behavioral responses in the distribution of labor income (Bergolo *et al.*, 2021).

are basically similar in the five cases, considering all individuals (panel or restricting the sample to those individuals whose income is reported by their employers/firms and who therefore do not file a tax return).

3. Correction Methods

In this section, we consider a hypothetical *true* distribution, $f_Y(y)$, and misreporting with a known shape. We can then derive the corresponding *distorted* distribution, $f_X(x)$, which suffers from average underreporting in high incomes, and study the impact of misreporting on standard correction methods.

3.1. *Simulation Design*

We consider the Singh–Maddala distribution, SM(2.257, 17393, 1.033), as the true distribution where a and q are shape parameters, b is a scale parameter, and $y > 0$.[10]These parameters are obtained by estimating a Singh–Maddala distribution from the Uruguayan linked data, combining survey data below and tax data above the median of the tax distribution. To mimic the underreporting obtained from the Uruguayan linked data, we assume that underreporting is patterned such that, on average, it increases above the median, as a piecewise linear model:

$$r(p) = \begin{cases} 1, & \text{if } p \leq 0.5, \\ 0.25 + 1.5p, & \text{if } 0.5 < p \leq 0.9, \\ -7.85 + 10.5p, & \text{if } p > 0.9, \end{cases} \tag{1}$$

where p is the proportion of income smaller than y in the true distribution, $p = F_Y(y)$, and $F_Y(y)$ is the CDF of the true distribution.[11]Under this design, on average, there is no underreporting below the median, and underreporting increases slowly above the median, until the 90th-quantile above which it increases sharply. More generally, underreporting can be defined with a function $r(p)$ such that $r(p) \geq 1$. Thus, we can obtain misreported incomes from the following relationship:

$$y = x \; r(p) \; \varepsilon, \quad \text{where } \varepsilon \sim N(1, \sigma^2). \tag{2}$$

A misreported income x is then obtained by dividing a true income y by a misreporting factor $r(p)\,\varepsilon$, which is on average equal to $r(p)$. The parameter σ measures the heterogeneity of misreporting rates of individuals with the same tax income. When $\sigma = 0$, individuals with the same tax income misreport exactly the same amount. We use $\sigma = 0.15$ to introduce some heterogeneity. Moreover, we restrict ε to be strictly positive, to ensure positive income $x > 0$.

Figure 3 shows the average misreporting rates, $r(p)$, and 1,000 true misreporting rates generated from the process described above, $y/x = r(p)\varepsilon$. We can see that

[10]The Singh–Maddala density function is equal to $f(y) = aqy^{a-1}/\{b^a[1 + (y/b)^a]^{1+q}\}$.

[11]The constant terms 0.25 and −7.6 are defined to have a continuous function at the knots 0.5 and 0.9.

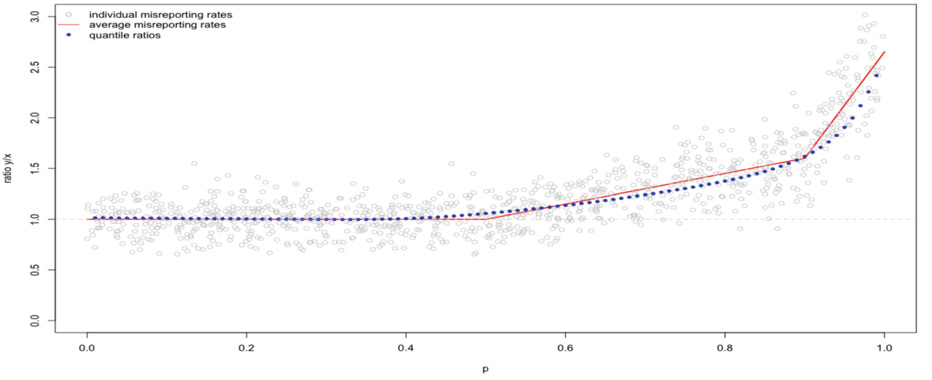

Figure 3. Simulation Design: Average Misreporting Rate Function $r(p)$, Quantile Ratios $Q_Y(p)/Q_X(p)$, and 1,000 Simulated Misreporting Rates, $y/x = r(p)\epsilon$. [Colour figure can be viewed at wileyonlinelibrary.com].

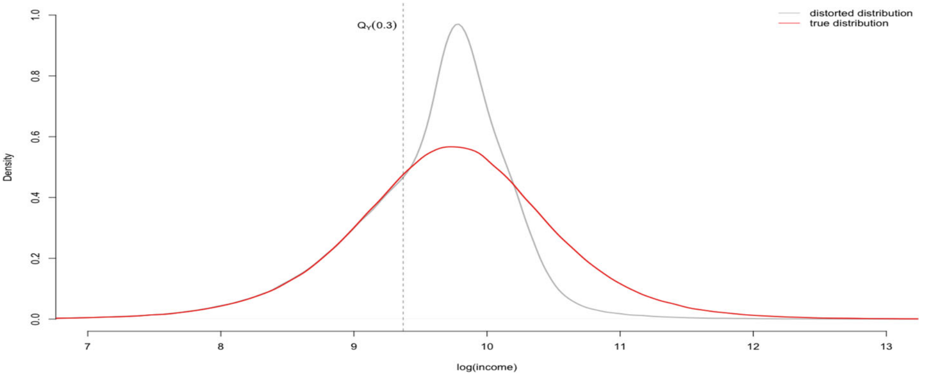

Figure 4. True Hypothetical Distribution (Black Line), and Distorted Distribution (Gray Line) That Suffers from Underreporting. [Colour figure can be viewed at wileyonlinelibrary.com].

the simulated data mimic the underreporting shape obtained at high incomes from linked data (see Figure 2).[12]

Figure 4 shows the density function of the true distribution (in logs) and a kernel density estimation of the distorted distribution (in logs), obtained from a sample of 1 million observations generated using the process described above. We can see that the distorted distribution, which suffers from average underreporting in high values, deviates significantly from the true distribution in the upper part of the distribution.

Table 2, rows 1 and 2 (`true`, `distorted`), shows several inequality indices computed from the true and distorted distributions. We can see that inequality is always smaller in the distorted distribution. In other words, inequality is downward

[12]As we do not have the analytical formula for the distorted distribution, we use a huge sample (1 million observations) to approximate it.

TABLE 2
Simulated Data: Inequality Measures Computed from the True and Distorted Distributions, and from Correction Methods with Several Tax-Quantile Thresholds *t*

t	Gini	MLD	Theil	Top 10 percent	Top 5 percent	Top 1 percent
True	0.436	0.335	0.385	0.339	0.230	0.093
Distorted	0.299	0.174	0.173	0.226	0.140	0.052
Replacing						
*q*90	0.424	0.324	0.412	0.377	0.257	0.104
*q*67.6	0.441	0.342	0.397	0.346	0.235	0.095
*q*50	0.437	0.338	0.387	0.340	0.231	0.093
*q*40	0.436	0.337	0.385	0.339	0.231	0.093
*q*30	0.436	0.337	0.385	0.339	0.231	0.093
*q*25	0.436	0.337	0.385	0.339	0.231	0.093
Reweighting						
*q*90	0.422	0.316	0.386	0.355	0.241	0.098
*q*67.6	0.416	0.307	0.358	0.330	0.225	0.091
*q*50	0.427	0.322	0.373	0.335	0.228	0.092
*q*40	0.435	0.335	0.384	0.339	0.230	0.093
*q*30	0.436	0.337	0.385	0.339	0.231	0.093
*q*25	0.436	0.337	0.385	0.339	0.230	0.093
BFM						
*q*90	0.421	0.316	0.385	0.355	0.241	0.098
*q*67.6	0.416	0.307	0.357	0.330	0.225	0.091
*q*50	0.427	0.322	0.372	0.335	0.228	0.092
*q*40	0.435	0.335	0.383	0.339	0.230	0.093
*q*30	0.436	0.337	0.384	0.339	0.231	0.093
*q*25	0.436	0.337	0.384	0.339	0.230	0.093

biased when underreporting occurs in high incomes, due to the fact that the distorted distribution differs from the true distribution, i.e., due to *non-sampling* errors.

Simulation experiments with different underreporting shapes have been investigated. For instance, underreporting mainly concentrated in the upper tail of the distribution is sometimes considered (Jenkins, 2017; Piketty *et al.*, 2019). In the Appendix, we report the results when increasing average underreporting occurs in the top 5 percent only (see Table A.2.2).

In general, we need external information to correct the problem of misreporting. In the following, we study the impact of underreporting at high incomes and the use of several correction methods, when externally reliable information is available in the upper part of the true distribution. In practice, the survey distribution is often considered to suffer from underreporting at high incomes, and the tax distribution is often considered to be more reliable external information for top incomes (Jenkins, 2017). The opposite is often considered for low incomes, with survey data more reliable than tax data (when available). Correction methods are then used to combine these two distributions.

3.2. *Replacing*

A first correction method consists of replacing misreported incomes above a threshold based on their corresponding quantiles in the true distribution. Let us consider a cumulative distribution function (CDF), F_X, which is continuous and strictly monotonically increasing. The quantile function is the inverse function of the CDF:

$$Q_X(p) = F^{-1}(p). \tag{3}$$

The correction is done by multiplying misreported incomes x by quantile ratios above a given threshold s:

$$z = \begin{cases} x, & \text{when } x \leq s, \\ x\, Q_Y(p)/Q_X(p), & \text{when } x > s, \end{cases} \quad \text{with } p = F_X(x). \tag{4}$$

This method serves to replace misreported incomes above s with their corresponding quantiles in the true distribution, since $Q_X(p) = x$.

The corrected distribution obtained by replacing, henceforth called *replaced* distribution, is defined as follows:

$$f_z(x) = \begin{cases} f_X(x), & \text{when } x \leq s, \\ 0, & \text{when } s < x \leq t, \\ f_Y(x), & \text{when } x > t, \end{cases}$$

where s and t are the $(100 - k)$th-quantile of the distorted and true distributions, respectively:

$$s = Q_X(1 - k/100) \quad \text{and} \quad t = Q_Y(1 - k/100). \tag{6}$$

The replaced distribution is then defined with the bottom $(100 - k)$ percent of the distorted distribution and the top k percent of the true distribution. It can also be obtained by combining misreported data below s and true data above t, i.e., by replacing the top k percent of the distorted distribution with the top k percent of the true distribution.

Theoretically, multiplying the misreported data by the quantile ratios, as defined in (4), is equivalent to replacing the top k percent of the distorted incomes by the top k percent of the true incomes. This is true when we consider the whole population an infinite number of individuals. However, when using finite samples, the results can differ significantly (see discussion in Section 4.1).

When microdata are available, we should combine the $(100 - k)$ percent lowest misreported data with the k percent highest true data. If the number of observations in the k percent highest misreported and true data are not the same, we must reweight to guarantee that the selected true data represent k percent of the combined sample.[13] Thus, we should apply weights equal to (n_z/n_x) to the selected misreported data, and equal to $(n_z/n_x)(m_x/m_y)$ to the selected true data, where n_x and n_z are the number of observations in the misreported and combined data, respectively, and m_x and m_y are the number of observations in the k percent highest misreported and true data, respectively.

[13]This reweighting procedure should not be confused with the reweighting method described in 3.3, which is designed to correct for underreporting or missing people.

From (5), we can see that, when there is no underreporting below the threshold s, it will be the case that $f_X(x) = f_Y(x)$ when $x \leq s$ and $s = t$, so the replaced distribution will be the true distribution. However, when underreporting occurs below s, the replaced distribution will deviate from the true distribution. Specifically, the $(100 - k)$th-quantile of the distorted and true distributions may differ, $s \neq t$, and the density will be equal to zero between these two values.

Finally, when the top k percent of the distorted distribution is replaced by the top k percent of the true distribution, additively decomposable inequality measures can be easily estimated from a non-overlapping decomposition, using a breakdown such as:

$$\begin{aligned} \text{Total inequality} &= \text{inequality of the smallest } (100-k) \text{ percent misreported data} \\ &+ \text{inequality of the highest } k \text{ percent true data} \\ &+ \text{between group inequality.} \end{aligned} \tag{7}$$

Decomposition formulas for the Gini and other inequality measures can be found in Alvaredo (2011) and Cowell (2011). Moreover, top v percent shares above t are defined as follows:

$$\mathrm{TS}_r(v) = \frac{(v/100)\mathbb{E}(y \geq Q_Y(1 - v/100))}{\mu_r} = \frac{\mu_Y}{\mu_r}\,\mathrm{TS}_Y(v), \quad \text{if } v \leq k, \tag{8}$$

where μ_r is the mean of the replaced distribution. When $v \leq k$, the top shares of the hybrid distribution are then equal to the top shares of the true distribution, rescaled by the mean ratio. Thus, when the mean of the replaced distribution is smaller (larger) than the mean of the true distribution, the top v percent shares with $v \leq k$ are biased upwards (downwards).

Figure 5(a) shows the replaced distribution (in logs) when we replace the top 10 percent of the distorted distribution with the top 10 percent of the true distribution. The threshold t is then equal to the 90th-quantile of the true distribution. A histogram is obtained from a sample of 1 million observations. We can see that the replaced distribution is similar to the distorted distribution at the bottom, and to the tax distribution in the upper part, but there are no values between the two 90th-quantiles of the distorted and true distributions.

Figure 5(b)–(e) shows the replaced distributions (in logs) when we replace the top 32.4 percent, 50 percent, 60 percent, and 70 percent of the distribution. The threshold t is then equal to, respectively, the 67.6th-, 50th-, 40th-, and 30th-quantiles of the true distribution. We can see that the replaced distribution gets closer to the true distribution as the threshold decreases. Let us define the optimal threshold as the value above which the true and distorted distributions start to differ, which is around the 30th-quantile (see Figure 4). An interesting feature of this correction method is that the replaced distribution deviates from the true distribution locally, i.e., between the optimal threshold and the selected threshold only.

Table 2, (`replacing`), shows inequality measures obtained from the replacing method with several different thresholds. We can see that the inequality measures are much closer to the true values than those obtained from the distorted distribution. Nevertheless, substantial differences remain for the 90th- and 67.6th-quantile

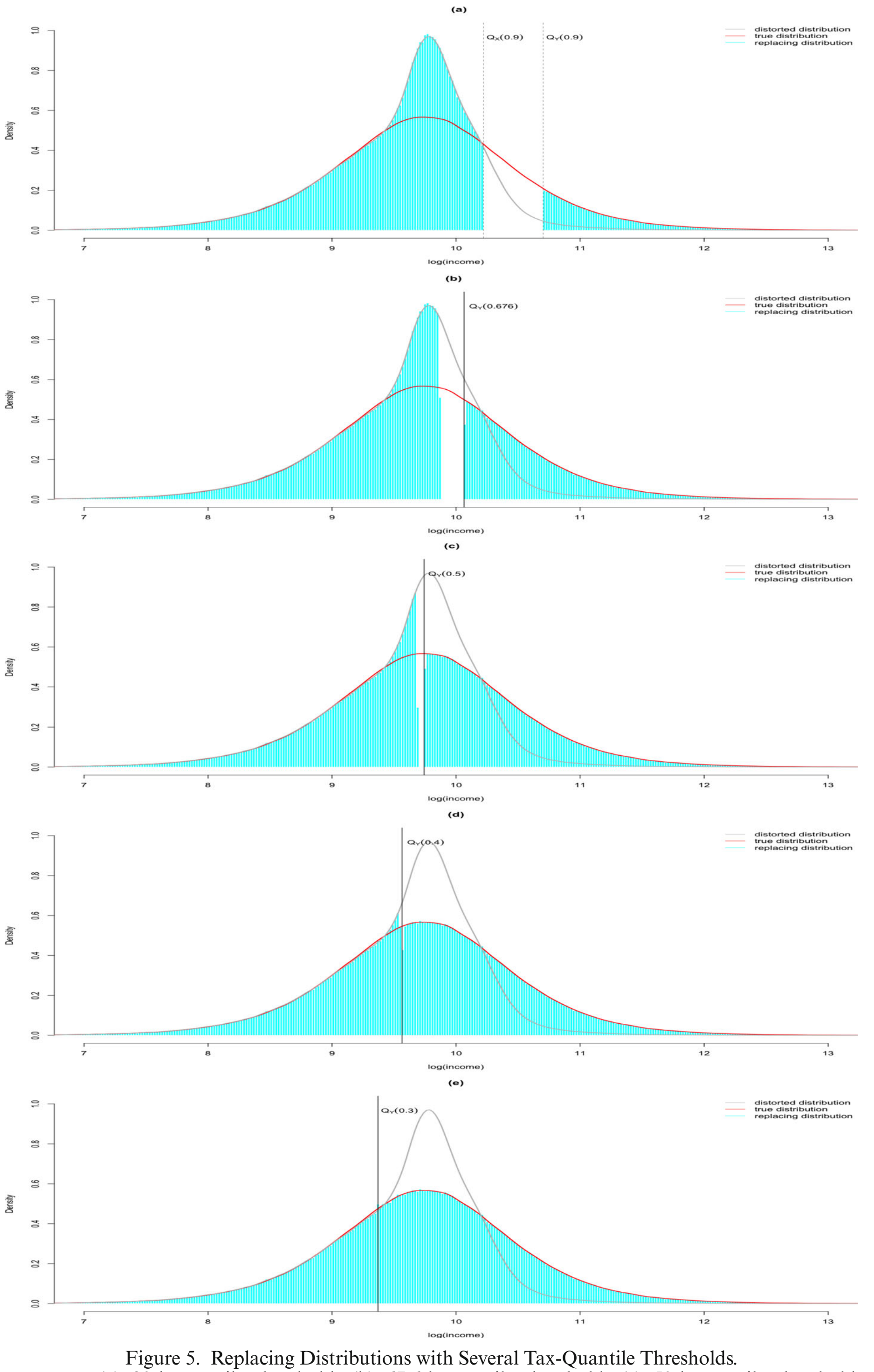

Figure 5. Replacing Distributions with Several Tax-Quantile Thresholds.

Notes: (a) 90th-quantile threshold, (b) 67.6th-quantile threshold, (c) 50th-quantile threshold, (d) 40th-quantile threshold, and (e) 30th-quantile threshold. [Colour figure can be viewed at wileyonlinelibrary.com].

thresholds, and the top 10 percent, 5 percent, and 1 percent shares are overestimated. On the contrary, the results are very good for the 50th-, 40th-, 30th-, and 25th-quantile thresholds.

This correction method is often used in empirical studies. Among others, see Burkhauser *et al.* (2016), Jenkins (2017), Nhlasny2017impact, Hlasny and Verme (2018), Piketty *et al.* (2019), and Chancel and Piketty (2019).

3.3. *Reweighting*

A second correction method consists of reweighting misreported data to recover the upper tail of the true distribution above a threshold, with the following weights:

$$w(x) = \begin{cases} \lambda, & \text{if } x \leq t, \\ f_Y(x)/f_X(x), & \text{if } x > t, \end{cases} \tag{9}$$

where λ is a constant defined to ensure that the density function integrates to one.

The corrected distribution obtained by reweighting, henceforth called *reweighted* distribution, is given by $w(x)f_X(x)$, that is:

$$f_w(x) = \begin{cases} \lambda f_X(x), & \text{if } x \leq t, \\ f_Y(x), & \text{if } x > t. \end{cases} \tag{10}$$

It is equal to the density of the true distribution above t and to the density of the distorted distribution rescaled by a factor λ below t.

This reweighted distribution can also be obtained by combining misreported data below a threshold t, and true data above t, with the following weights:

$$w'(z) = \begin{cases} (100 - l)/(100 - m), & \text{if } z \leq t, \\ l/m, & \text{if } z > t, \end{cases} \tag{11}$$

where $m = 100l/(100 - k + l)$, and l and k are implicitly defined from the quantile functions:

$$t = Q_X(1 - k/100) = Q_Y(1 - l/100). \tag{12}$$

The underlying distribution is then defined using the bottom $(100 - k)$ percent of the distorted distribution, and the top l percent of the true distribution. When $k \neq l$, there is an implicit reweighting: misreported and true data correspond, respectively, to the bottom $100 - m$ percent and to the top m percent of the hybrid distribution. The role of the weights is to increase the density above t, such that the top l percent of the reweighted distribution corresponds to the top l percent of the true distribution, and to decrease the density below t to compensate. Finally, combining misreported data below t and true data above t with the weights defined in (11) leads to the density function in (10), with $\lambda = (100 - l)/(100 - m)$.

Theoretically, using misreported data with the weights defined in (9) is equivalent to combining misreported data below t and true data above t with the weights defined in (11). This is true when we consider the whole population. However, in a finite sample, density ratios are difficult to accurately estimate for high incomes, where densities are close to zero, and the weights in (9) may therefore be unreliable. In practice, the results can differ significantly between the two approaches (see discussion in Section 4.1).

Blanchet *et al.* (2019), denoted BFM hereafter, proposed a correction method that is used in the *World Inequality Database*; see Alvaredo *et al.* (2020). This method involves a first step, in which misreported (survey) incomes are used with the weights defined in (9). To correct, among other things, problems induced by the poor estimation of density ratios, a second step is performed in which misreported (survey) observations above a threshold are duplicated several times and replaced by observations with equivalent rank and weight in the true (tax) distribution.

In the end, the numerical results are quite similar to those obtained by combining misreported (survey) data below t and true (tax) data above t with the weights defined in (11).[14]

From (11) and (12), we can see that when there is no underreporting below t, we have $k = l = m$, $w(x) = 1$, and the reweighted distribution is the true distribution. However, when underreporting occurs below the threshold t, we have $k \neq l$, and the reweighted distribution deviates from the true distribution below t.

Finally, when misreported data are used below a threshold t, and true data are used above t, additively decomposable inequality measures can be easily estimated from a non-overlapping weighted decomposition, using a breakdown such as the following one:

$$\begin{aligned} \text{Total inequality} &= \text{inequality of the misreported data below } t \\ &\quad + \text{inequality of the true data above } t \\ &\quad + \text{between group inequality} \end{aligned} \tag{13}$$

with the weights defined in (11). Because the weights are constant in each group, weighted decomposition formulas for inequality measures with the property of scale independence are similar to unweighted decomposition formulas, where the share of the misreported data below t is equal to $1 - l/100$, the share of the true data above t is equal to $l/100$, and the overall mean μ_w is the weighted mean of the two groups:

$$\mu_w = (1 - l/100)\, \mathbb{E}(x|x \leq t) + (l/100)\, \mathbb{E}(y|y > t). \tag{14}$$

Moreover, the top v percent shares above the threshold t in (12) are defined as follows:

$$\text{TS}_w(v) = \frac{(v/100)\mathbb{E}(y \geq Q_Y(1 - s/100))}{\mu_w} = \frac{\mu_Y}{\mu_w}\, \text{TS}_Y(v), \quad \text{if } \; v \leq l. \tag{15}$$

They are equal to the top shares obtained from the true data, rescaled by the mean ratio. Thus, when the mean of the reweighted distribution is smaller (larger) than the

[14]There is an issue concerning the threshold selection embedded in the BFM method. See Section 3.4 for a discussion on this.

mean of the true distribution, the top v percent shares with $v \leq l$ are biased upwards (downwards).

Figure 6(a) shows the reweighted distribution (in logs), combining misreported data below t and true data above t, when the threshold is the 90th-quantile of the true distribution. A histogram is obtained from a sample of 1 million observations. We can see that the reweighting distribution is similar to the true distribution above the threshold, and it is similar to the distorted distribution pushed downwards below the threshold.

Figure 6(b)–(e) shows the reweighted distributions (in logs), combining misreported data below t and true data above t, when t is, respectively, the 67.6th-, 50th-, 40th-, and 30th-quantile of the true distribution. We can see that the reweighted distribution gets closer to the true distribution as the threshold decreases. A specific feature of this correction method is that the reweighted distribution deviates from the true distribution everywhere below the selected threshold.

Table 2, (`reweighting`), shows the inequality measures obtained from the reweighting method with several thresholds. We can see that the inequality measures are much closer to the true values than those obtained from the distorted distribution. Nevertheless, substantial differences remain when using the 90th-, 67.6th-, and 50th-quantile thresholds. On the contrary, the results are very good when using the 40th-, 30th-, and 25th-quantile thresholds. Table 2, (`BFM`), shows the inequality measures obtained with the BFM method. The results are identical to the reweighting method, which combines survey data below the threshold and tax data above with the weights defined in (10).

Compared to the replacing method, which returns values for inequality measures close to the true values more quickly as the threshold decreases, the reweighting method requires a lower threshold to obtain analogous results. This comes from the fact that the replacing distribution deviates from the true distribution locally (between the optimal threshold and the selected threshold only), while the reweighted distribution deviates from the true distribution more globally (everywhere below the selected threshold). Thus, when the selected threshold is not very far above the optimal threshold, the replacing distribution deviates from the true distribution in a quite narrow interval.

The reweighting method has been used in Anand and Segal (2017), Hlasny and Verme (2017), Burkhauser *et al.* (2018), Hlasny and Verme (2018), Campos-Vazquez and Lustig (2020), and Department for Work and Pensions (2015).

3.4. *Threshold Selection*

When there is, on average, no underreporting below the selected threshold, the replacing and reweighting methods are similar. They are based on a distribution that is the true distribution, and thus they effectively correct the problem of misreporting. However, when underreporting occurs below the threshold, the previous subsections show that the results may be biased. The choice of threshold is therefore a key issue.

From Figure 4, we can see that the true and distorted distributions begin to deviate around the 30th-quantile. This suggests that, in our simulation design,

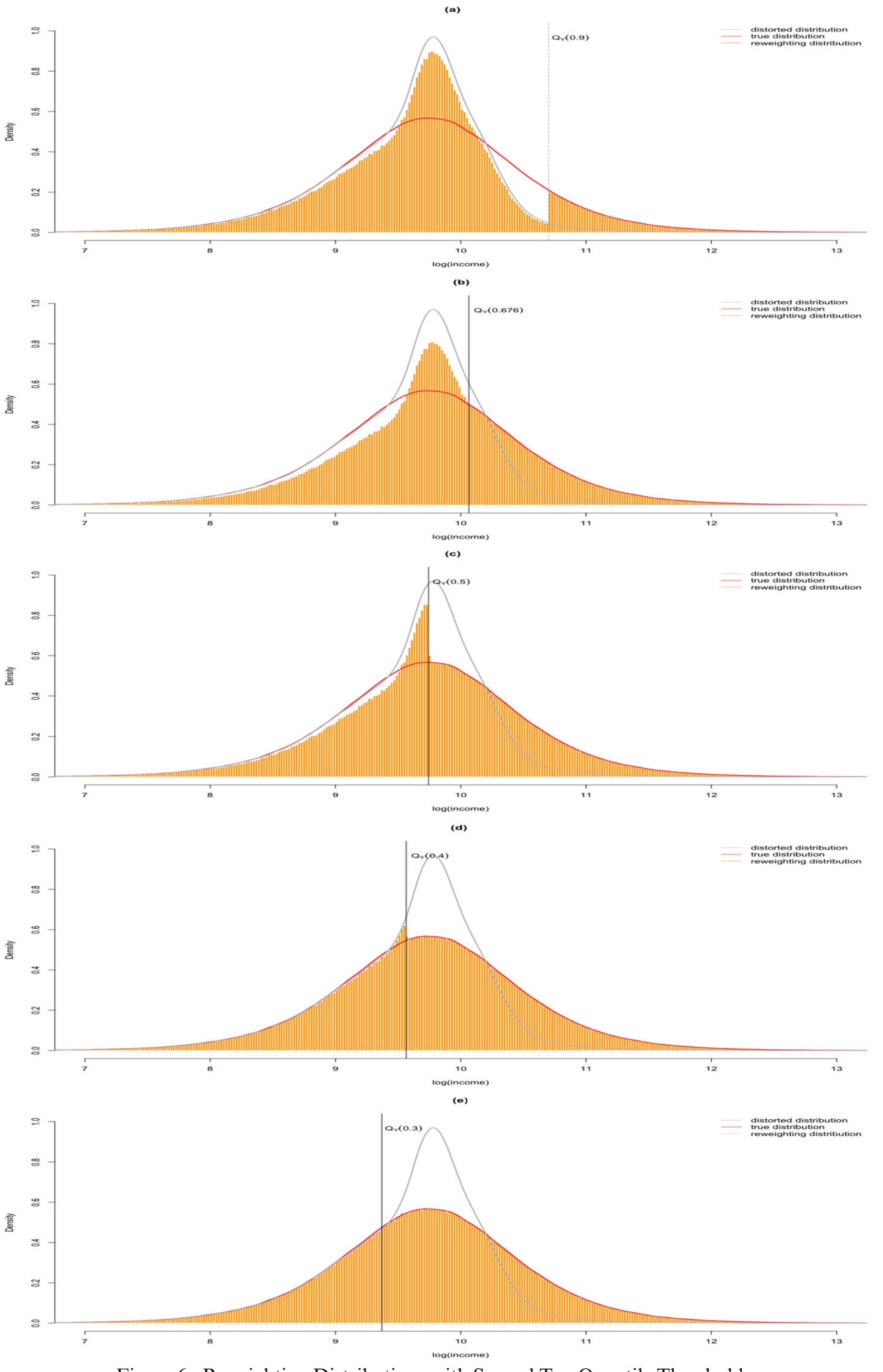

Figure 6. Reweighting Distributions with Several Tax-Quantile Thresholds.

Notes: (a) 90th-quantile threshold, (b) 67.6th-quantile threshold, (c) 50th-quantile threshold, (d) 40th-quantile threshold, and (e) 30th-quantile threshold. [Colour figure can be viewed at wileyonlinelibrary.com].

the optimal threshold is around 30th-quantile. It is below the median, which may be surprising because there is no *average* misreporting below the median; see (1). This is due to the heterogeneity of misreporting behaviors, defined by $\sigma > 0$ in (1). Indeed, above the median, some individuals overreport their income, which is then replaced by a lower income, which may be below the median. The distribution below the median is then affected by *individual* misreporting. It follows that the optimal threshold may be smaller than the value above which average misreporting rates increase.

In practice, the threshold can be selected a priori, as the 80th, 90th, 95th, or 99th quantile of the distribution (Burkhauser *et al.*, 2016; Chancel and Piketty, 2019; Piketty *et al.*, 2019). A less arbitrary method consists of selecting the threshold based on the quantile ratio function:

$$\text{(16)} \qquad \text{select} \quad t = \max(Q_X(p)) \quad \text{such that} \quad \frac{Q_Y(p)}{Q_X(p)} = 1.$$

As long as the true and distorted distributions are identical at the bottom (below the optimal threshold), they share similar quantiles. This method is then designed to detect the value above which the two distributions differ when underreporting occurs above a threshold and when there are no other measurement errors.

To illustrate, Figure 3 shows the quantile ratios in our simulation design. We can see that the quantile ratios begin to deviate from 1 below the median. From this figure, we would select a threshold at around the 40th-quantile of the tax distribution. The replacing and reweighting methods provide inequality measures very close to the true values with this threshold (see Table 2).

Another threshold selection, proposed by Blanchet *et al.* (2019), is as follows:

$$\text{(17)} \qquad \text{select} \quad t = \max(z) \quad \text{such that} \quad \frac{F_X(z)}{F_Y(z)} = \frac{f_X(z)}{f_Y(z)}.$$

This method is defined to ensure the continuity of the reweighting distribution in the upper tail. However, it is not designed to identify when the true and distorted distributions start to differ, and it often selects a threshold that is too high. Indeed, the selected threshold is close to the highest of possible crossing-points between both densities. If the two distributions deviate above this crossing-point, by definition, they will also deviate at least by the same magnitude/area below this crossing-point (they are both density functions). Therefore, using this threshold and reweighting, it is assumed that the deviation below the threshold is equally distributed over the distribution below the threshold. This implies that underreporting has an impact across the entire distribution, and this impact is precisely known below about the highest crossing-points. If underreporting differs from this specific design, a lower threshold will perform better. Moreover, as density ratios are difficult to estimate in finite samples, the results can be very erratic. For instance, the merging points vary significantly across the years in empirical applications with high-quality tax data such as that from Norway and France in Blanchet *et al.* (2019), ranging from the 60th- to the 99th-quantiles over the three years from 2007 to 2009.

Figure 7 shows the threshold selection obtained with (17) on our simulated data. The density ratio function, $\theta(x) = f_X(x)/f_Y(x)$, and the CDF ratio function,

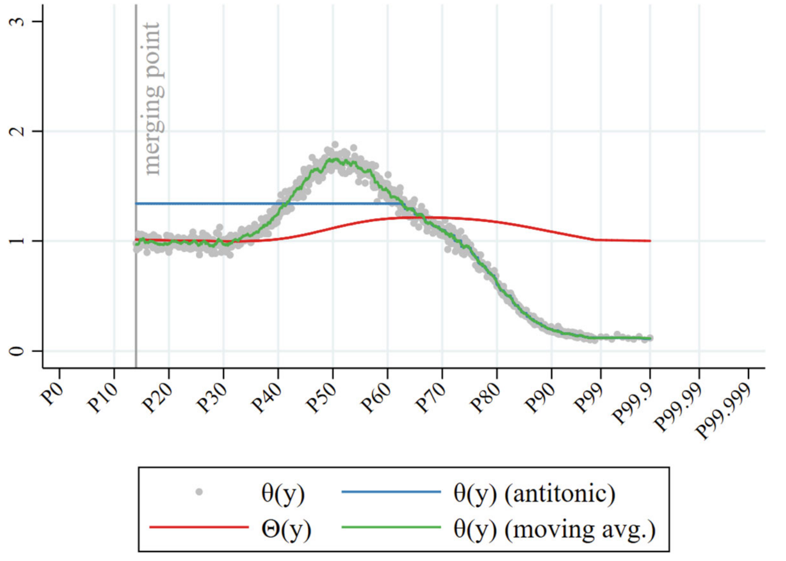

Figure 7. Simulated Data: Density Ratio Function, $\theta(y) = f_X(y)/f_Y(y)$, and CDF Ratio Function, $\Theta(y) = F_X(y)/F_Y(y)$, Obtained with BFM Method. [Colour figure can be viewed at wileyonlinelibrary.com].

$\Theta(x) = F_X(x)/F_Y(x)$, are plotted. We can see that the moving averages θ and Θ are close to 1 until the 30th-quantile, above which they start to deviate, which corresponds to the special case (16) and to the optimal threshold detected in Figure 4. However, the threshold selected using (17) is much higher; it is obtained when θ and Θ cross at the 67.6th-quantile of the tax distribution. Figure 6(b) shows the reweighting distribution obtained with this threshold. We can see that the distribution is continuous, but it deviates significantly from the true distribution everywhere below the selected threshold. Moreover, all correction methods applied with this threshold provide inequality measures substantially different from their true values (see Table 2).

4. What to Do in Practice?

In practice, tax data are often considered more (less) reliable than survey data for high (low) incomes. Correction methods are then used to combine these two distributions. In such cases, what correction approach should be recommended in practice, based on the previous results? And what would the results look like if we apply the recommended strategy to Uruguayan linked data?

4.1. *Lessons from the Simulation Results*

First, we should stress that our simulation results are robust to many different underreporting patterns. In particular, we find similar results when underreporting is mainly concentrated at the top of the distribution (see additional results in the Appendix).[15]

The simulation results are focused on misreporting issues. They are based on population distributions, so they are not subject to finite sample issues.[16]Two main lessons can be drawn from the results:

- The performance of the correction methods is quite sensitive to the threshold selection, which appears to be a key issue.
- The replacing method is less sensitive than the reweighting method to the threshold selection.

In practice, the optimal threshold is unknown. Moreover, it is more difficult to select an appropriate threshold if tax data are not reliable for low incomes, as quantile ratios may then differ from unity with thresholds lower than the optimal threshold. Therefore, we cannot rely as much on the method based on quantile ratios in (16) and, in general, it is not easy to select the threshold in practice. Moreover, tax records are not without problems in the upper tail because of tax evasion and avoidance (Atkinson, 2007).

Without an a priori decision on the value above (below) which tax (survey) data are reliable, the threshold selection based on quantile ratios in (16) may be used as a starting point in practice, and in addition, results may be presented with several thresholds to check robustness/sensitivity to the threshold.[17]

The replacing and reweighting methods can each be implemented in two different ways: (1) by correcting/reweighting misreported (survey) data, or (2) by combining misreported (survey) and (true) tax data with/without weights; see Sections 3.2 and 3.3. At the population level, the two approaches are equivalent, but they can differ significantly in finite samples. Overall, approach (2), which combines misreported (survey) and true (tax) data, should be preferred in practice for the following reasons:

- The reweighting method based on misreported data with the weights defined in (9) may result in poor performance, because density ratios are difficult to estimate accurately in finite samples, especially in the upper tail.
- Misreporting is not the only reason to believe that surveys do not capture top incomes well. For instance, top-coding or censoring may be imposed by the data provider for reasons of confidentiality. Furthermore, some portion of the sampled population may not respond to the survey (item nonresponse), or may be difficult to reach easily (unit nonresponse). These additional reasons are all related to missing data, which is another major issue in surveys.[18]They

[15]Table A.2.2 shows the results with underreporting increasing linearly above the 95th-quantile, i.e., when (1) is replaced by $r(p) = 20p - 18$ if $p > 0.95$, otherwise $r(p) = 1$, and $\sigma = 0.05$ in (2).

[16]In fact, the population distributions are approximated using huge samples of 1 million observations.

[17]Note that the quantile functions may not cross if the misreported distribution first-order stochastic dominates the true distribution, so the selected threshold from (16) may be infeasible.

[18]Lustig (2019) uses *missing rich* as a catch-all term to refer to the causes of misreporting and missing data affecting the upper tail of survey distributions. These problems are also known as survey *undercoverage* of top incomes (Jenkins, 2017; Burkhauser *et al.*, 2018).

can be partially overcome by using external tax data in the upper tail, not by correcting/reweighting misreported survey data only.

Therefore, we recommend implementing the replacing method by combining the bottom $(100 - k)$ percent of the distorted (survey) distribution and the top k percent of the true (tax) distribution, and the reweighting method by combining misreported (survey) data below a threshold and true (tax) data above it, with the weights defined in (11).

Both the replacing and reweighting methods allow recovering the upper tail of the true distribution. The main difference between the two methods is that replacing leaves unchanged the bottom part of the misreported distribution, while reweighting modifies it by a scale factor. Thus, at least for the underreporting pattern in our simulated distorted distribution, replacing is more appropriate when average misreporting occurs in the upper part of the distribution, as illustrated in our simulation results. On the contrary, reweighting could be more appropriate when missing data occur in the upper part of the distribution, because the bottom part of the misreported distribution has to be adjusted when additional data are included at the top. It is possible to combine both the replacing and reweighting methods to correct misreporting and missing data issues; however, two different thresholds may be required, and there is no simple way to select them without external information. Again, the selection of the threshold(s) is a key issue.

Finally, unlike the previous section, which relies on population distributions, we also have to consider *sampling bias* in finite sample. In particular, it is well known that the empirical distribution function, or EDF, does not accurately capture the upper tail of heavy-tailed distributions, due to sparseness. This is still true with tax data, but at a much higher level than with survey data. A Pareto distribution is often fitted to the upper tail of income and wealth distributions to reduce sampling bias errors. An interpolation method with a GPD distribution adjusted to the top can be used with tabulated data (Blanchet *et al.*, 2017). The EDF with a Pareto distribution fitted to the top can be used with microdata (Charpentier and Flachaire, 2019).

The price paid to correct misreporting and missing data problems with these methods is that the covariates are lost, unless we make some strong assumptions. Indeed, when we multiply survey incomes by quantile ratios (replacing), or when we reweight survey incomes by density ratios (reweighting), we can keep covariates only if one of the following conditions holds: (1) misreporting preserves individual rankings in the income distribution or (2) individual rankings in the income distribution do not depend on the covariates. Indeed, with replacing, we cannot assume that survey income and the corresponding corrected income belong to the same individual, because quantile ratios do not measure individual misreporting, except if (1) holds. Therefore, we cannot transfer the covariates of an individual with a given survey income to the individual with the corresponding corrected income, except if (1) or (2) holds. With reweighting, if individuals are reranked at lower/higher levels in the survey, due to misreporting, we cannot use their covariates because their true positions in the income distribution are unknown, except if (1) or (2) holds. To illustrate, let us consider an example that is extreme, but helps drive the point home. If the richest individual in the true distribution underreports his income in the survey in such a way that he is ranked as the poorest individual in the survey, it would make no sense to assign his covariates to the lowest survey income.

Finally, as soon as misreporting implies individual rerankings in the income distribution, the link to individuals and therefore the link to covariates are lost, unless we use linked data or we make unrealistic assumptions such as (1) or (2). With our linked data, we find that reranking can occur to a significant extent when one goes from the survey to the tax distribution: the same individual switches ranks, sometimes by a lot. Inequality measures, as they are anonymous, are not affected by reranking.

4.2. *Application with Uruguayan Linked Data*

In this section, we apply several correction methods to the Uruguayan linked survey and tax data. Tax data are considered more reliable than survey data for high incomes, but they are known to be unreliable below the minimum wage. These linked data have shown evidence of average underreporting at high (low) incomes for the survey (tax) data; see Figure 2. With underreporting of high incomes in survey data, inequality measures from the survey would be biased. With underreporting of low incomes in tax returns, inequality would also be biased in tax data. Using correction methods, we seek to correct these biases by combining survey and tax data.

Figure 8 shows the quantile ratios computed with Uruguayan survey and tax data. We can see that the quantile ratios are smaller (greater) than 1 below (above) the 50th-quantile. Thus, we should select a threshold around the median. This is consistent with Figure 2, which shows the evidence of average underreporting above the median in the Uruguayan survey data. Moreover, the BFM threshold, which ensures continuity of the reweighting distribution in the upper tail, as defined in (17), is selected at the 72th-quantile of the tax distribution, which, as expected based on the previous discussion, is too high.

Table 3 shows the inequality measures computed from the Uruguayan data: using the survey and tax data, and using the correction methods with several different thresholds. The number of observations is equal to $n = 1461$. After merging the two data sets, a Pareto distribution is fitted to the top 5 percent of each distribution for the replacing and reweighting methods.[19]The main results can be summarized as follows:

- Correction methods provide inequality values that are larger than those from the survey and smaller than those from tax data. It suggests that the survey data underestimate and the tax data overestimate inequality measures.
- Reweighting and BFM provide very similar results when one chooses the same threshold for both, but they are not exactly the same in finite samples. BFM is based on an interpolation method with a Generalized Pareto distribution adjusted at the extreme top, whereas our implementation of reweighting is based on a Pareto distribution fitted to the top 5 percent.
- Replacing, reweighting, and BFM have the most similar results when using the 50th-quantile threshold, which is the threshold selected with quantile ratios, and above which the linked data provide evidence of underreporting (see Figure 2).

[19]The number of observations is too small to fit a Pareto distribution in the top 1 percent or higher.

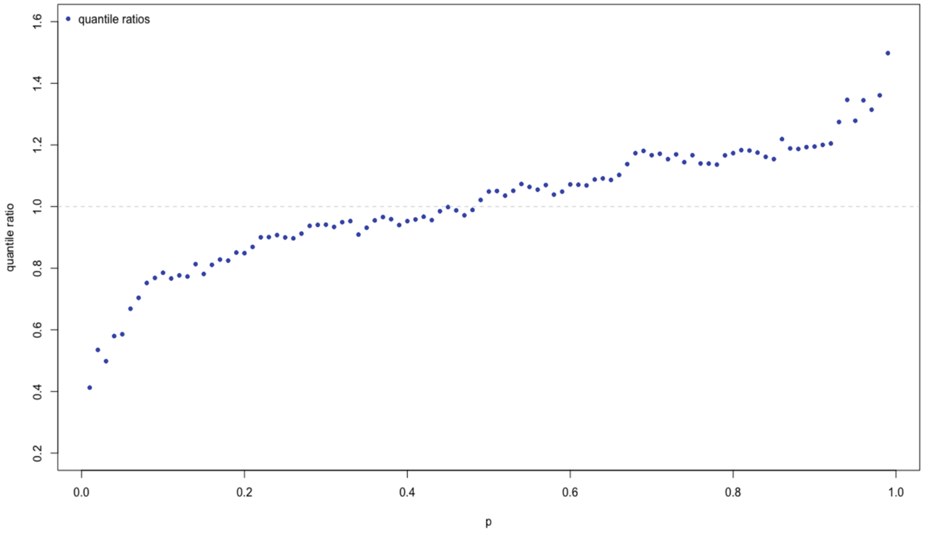

Figure 8. Uruguayan Linked Data: Quantile Ratios, $\hat{Q}_Y(p)/\hat{Q}_X(p)$. [Colour figure can be viewed at wileyonlinelibrary.com].

TABLE 3
Uruguayan Linked Data: Inequality Measures Computed from the Tax and Survey Samples, and from Correction Methods with Several Tax-Quantile Thresholds *t*

t	Gini	MLD	Theil	Top 10 percent	Top 5 percent	Top 1 percent
Tax	0.472	0.423	0.448	0.359	0.247	0.102
Survey	0.382	0.254	0.272	0.300	0.192	0.068
Replacing						
*q*90	0.440	0.336	0.419	0.373	0.253	0.104
*q*72	0.446	0.343	0.414	0.355	0.244	0.100
*q*60	0.447	0.346	0.412	0.353	0.243	0.100
*q*50	0.447	0.346	0.410	0.350	0.241	0.099
*q*40	0.448	0.347	0.412	0.351	0.242	0.099
*q*30	0.450	0.350	0.414	0.351	0.242	0.099
Reweighting						
*q*90	0.442	0.338	0.409	0.358	0.246	0.100
*q*72	0.435	0.330	0.391	0.344	0.237	0.097
*q*60	0.441	0.337	0.400	0.348	0.239	0.098
*q*50	0.443	0.340	0.402	0.349	0.239	0.097
*q*40	0.452	0.354	0.414	0.352	0.242	0.098
*q*30	0.458	0.365	0.422	0.354	0.243	0.099
BFM						
*q*90	0.443	0.341	0.408	0.362	0.250	0.102
*q*72	0.435	0.332	0.387	0.347	0.239	0.097
*q*60	0.441	0.340	0.397	0.350	0.242	0.098
*q*50	0.444	0.344	0.400	0.352	0.240	0.099
*q*40	0.452	0.357	0.412	0.355	0.242	0.102
*q*30	0.458	0.369	0.421	0.358	0.245	0.104

- Replacing provides very stable results. For instance, numerical differences do not exceed 0.003 with thresholds at the 60th-, 50th-, and 40th-quantiles, and they do not exceed 0.007 with thresholds at the 72th-, 60th-, 50th-, 40th-, and 30th-quantiles.

Overall, the empirical results are quite similar to our simulation results. The replacing method provides more stable results than the other methods, using different thresholds.

5. Conclusions

Household surveys suffer from sampling and non-sampling errors, and these errors result in biased inequality measures. We compared correction methods to address one type of measurement error: underreporting of income in the upper tail. Two main approaches have been used in the literature to correct survey upper-tail errors, including underreporting: replacing and reweighting. The key difference between the two methods is whether, for any chosen threshold above which one contends there is underreporting, the total weights of the upper-tail and the rest of the distribution remain unchanged (replacing) or not (reweighting). Both correction approaches rely on appropriate selection of the threshold beyond which survey data tend to underreport income. The biggest challenge in applying correction methods is that—as the true distribution is unknown—the true threshold above which underreporting occurs is also unknown.

To assess the implications of alternative correction methods on inequality estimates, we relied on simulation. We considered a true distribution and constructed a distorted distribution that features underreporting of income in the upper tail. The pattern of underreporting in our simulation mimics the pattern observed in a novel data set that links individuals from Uruguay's household survey to their tax returns.

As anticipated, our simulations show that threshold selection plays a key role. If the threshold is not correctly chosen, inequality measures may be significantly biased. An interesting finding is that the replacing method is less sensitive to the choice of the threshold. By "less sensitive," we mean that inequality measures vary less when different thresholds are selected. If inequality measures are quite similar when using different thresholds, this suggests that selecting a threshold slightly different from the correct one has no/little impact on the results. In addition to the simulations, we explored how to approach the threshold selection challenge in practice using the linked data for Uruguay. Our results are analogous to our simulation exercise: with replacing, the inequality measures are less sensitive to the threshold selection.

Our exercises seem to imply that, to address underreporting in the upper tail, the replacing method may be preferable given that the inequality measures it generates are less sensitive to threshold selection. This is not necessarily the case if one wants to correct for other errors such as missing data. As household surveys are likely to suffer from multiple errors, one should test the sensitivity of inequality measures to alternative thresholds. If the results turn out to be very sensitive to the choice of method and threshold, one should report a range of values rather than choosing a single one.

The results presented in this article should not be considered a general assessment of the two approaches. "Reweighting" and "replacing" are two broad classes of methods, among which some specific applications of replacing and reweighting may outperform others. Therefore, one particular replacing method may outperform one particular reweighting method but this ranking may not hold for all replacing/reweighting methods.

References

Abowd, J. M. and M. H. Stinson, "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative data," *Review of Economics and Statistics*, 95(5), 1451–67, 2013.

Adriaans, J., P. Valet and S. Liebig, "Comparing Administrative and Survey Data: Is Information on Education from Administrative Records of the German Institute for Employment Research Consistent with Survey Self-Reports?" *Quality & Quantity*, 54(1), 3–25, 2020.

Altimir, O., "Income Distribution Statistics in Latin America and Their Reliability," *Review of Income and Wealth*, 33(2), 111–55, 1987.

Alvaredo, F., "A Note on the Relationship Between Top Income Shares And the Gini Coefficient," *Economics Letters*, 110, 274–7, 2011.

Alvaredo, F., A. Atkinson, T. Blanchet, L. Chancel, L. Bauluz, M. Fisher-Post, I. Flores, B. Garbinti, J. Goupille-Lebret, C. Martinez-Toledano, M. Morgan, T. Neef, T. Piketty, A.-S. Robilliard, E. Saez, L. Yang and G. Zucman, "Distributional National Accounts Guidelines 2020: Methods and Concepts Used in the World Inequality Database," Wid.world Working Paper, 2020.

Anand, S. and P. Segal, "Who Are the Global Top 1%?" *World Development*, 95, 111–26, 2017.

Atkinson, A. B., "Measuring Top Incomes: Methodological Issues," in A. B. Atkinson and T. Piketty (eds), *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*, Vol 1, Oxford University Press, New York, 18–42, 2007.

Bergolo, M., G. Burdin, M. De Rosa, M. Giaccobasso and M. Leites, "Digging into the Channels of Bunching: Evidence from the Uruguayan Income Tax," *The Economic Journal*, 131(639), 2726–62, 2021.

Biemer, P. P. and S. L. Christ, "Weighting Survey Data," in E. D. de Leeuw, J. J. Hox and D. A. Dillman (eds), *International Handbook of Survey Methodology*, Routledge, New York, 317–41, 2008.

Blanchet, T., I. Flores and M. Morgan, "The Weight of the Rich: Improving Surveys Using Tax Data," Wid.world Working Paper 2018/12, 2019.

Blanchet, T., J. Fournier and T. Piketty, "Generalized Pareto Curves: Theory and Applications," Wid.world Working Paper 2017/3, 2017.

Bollinger, C. R., B. T. Hirsch, C. M. Hokayem and J. P. Ziliak, "Trouble in the Tails? What We Know About Earnings Nonresponse 30 Years After Lillard, Smith, and Welch," *Journal of Political Economy*, 127(5), 2143–85, 2019.

Brzezinski, M., "Do Wealth Distributions Follow Power Laws? Evidence from 'Rich Lists'," *Physica A: Statistical Mechanics and its Applications*, 406, 155–62, 2014.

Burdín, G., F. Esponda and A. Vigorito, "Inequality and Top Incomes in Uruguay: A Comparison Between Household Surveys and Income Tax Micro-Data," Commitment to Equity Working Paper 21, 2014.

Burkhauser, R. V., N. Hérault, S. Jenkins and R. Wilkins, "What Has Been Happening to UK Income Inequality Since the Mid-1990s? Answers from Reconciled and Combined Household Survey and Tax Return Data," NBER Working Paper 21991, 2016.

Burkhauser, R. V., N. Hérault, S. Jenkins and R. Wilkins, "Survey Undercoverage of top Incomes and Estimation of Inequality: What Is the Role of the UK's SPI Adjustment?" *Fiscal Studies*, 39, 213–40, 2018.

Campos-Vazquez, R. and N. Lustig, "Labour Income Inequality in Mexico: Puzzles Solved and Unsolved," *Journal of Economic and Social Measurement*, 44, 203–19, 2020.

Carrasco, P., A. Cichevski and I. Perazzo, "Evolución Reciente de las Principales Variables del Mercado Laboral Uruguayo," Serie Documentos de Trabajo 9/18, 2018.

Chancel, T. and T. Piketty, "Indian Income Inequality, 1922–2015: From British Raj to Billionaire Raj?" *Review of Income and Wealth*, 65, S33–62, 2019.

Charpentier, A. and E. Flachaire, "Pareto Models for Top Incomes," Working Paper hal-02145024, 2019.

Cowell, F. A., *Measuring Inequality*, 3rd ed., Oxford University Press, Oxford, 2011.
Cowell, F. A. and E. Flachaire, "Income Distribution and Inequality Measurement: The Problem of Extreme Values," *Journal of Econometrics*, 141(2), 1044–72, 2007.
Cowell, F. A. and E. Flachaire, "Statistical Methods for Distributional Analysis," in A. B. Atkinson, and F. Bourguignon (eds), *Handbook of Income Distribution*, Vol 2, Elsevier, North Holland, 359–465, 2015.
Deaton, A., "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)," *Review of Economics and Statistics*, 87, 1–19, 2005.
Department for Work and Pensions, *Households Below Average Income An Analysis of the Income Distribution 1994/95–2013/14*, Department for Work and Pensions, London, 2015.
Fisher, J. D., D. S. Johnson, T. M. Smeeding and J. P. Thompson, "Inequality in 3-D: INCOME, Consumption, and Wealth," *Review of Income and Wealth*, 68(1), 16–42, 2022.
Gottschalk, P. and M. Huynh, "Are Earnings Inequality and mobility overstated? The Impact of Nonclassical Measurement Error," *The Review of Economics and Statistics*, 92(2), 302–15, 2010.
Groves, R. M. and L. Lyberg, "Total Survey Error: Past, Present, and Future," *Public Opinion Quarterly*, 74(5), 849–79, 2010.
Higgins, S., N. Lustig and A. Vigorito, "The Rich Underreport Their Income: Assessing Biases in Inequality Estimates and Correction Methods Using Linked Survey and Tax Data," Technical Report, Tulane University, Department of Economics, 2018.
Hlasny, V. and P. Verme, "The Impact of Top Incomes Biases on the meAsurement of Inequality in the United States," *Oxford Bulletin of Economics and Statistics*, 84(4), 749–88, 2017.
Hlasny, V. and P. Verme, "Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods Using the EU SILC Data," *Econometrics*, 6(2), 30, 2018.
Hyslop, D. R. and W. Townsend, "Earnings Dynamics and Measurement Error in Matched Survey and Administrative Data," *Journal of Business & Economic Statistics*, 38(2), 457–69, 2020.
Instituto Nacional de Estadística, "Encuesta de Nutrición, Desarrollo Infantil y Salud (ENDIS)," 2013.
Instituto Nacional de Estadística, "Encuestas Continuas de Hogares," 2018.
Jenkins, S. P., "Pareto Models, Top Incomes and Recent Trends in UK Income Inequality," *Economica*, 84, 261–89, 2017.
Jenkins, S. P. and F. Rios-Avila, "Modelling Errors in Survey and Administrative Data on Employment Earnings: Sensitivity to the Fraction Assumed to Have Error-Free Earnings," *Economics Letters*, 192, 109253, 2020.
Kapteyn, A. and J. Y. Ypma, "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data," *Journal of Labor Economics*, 25(3), 513–51, 2007.
Lustig, N., "The Missing Rich in Household Surveys: Causes and Correction Approaches," CEQ Working Paper 75, 2019.
Meyer, B. D. and N. Mittag, "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net," *American Economic Journal: Applied Economics*, 11(2), 176–204, 2019.
Piketty, T., E. Saez and G. Zucman, "Distributional National Accounts: Methods and Estimates for the United States," *The Quarterly Journal of Economics*, 133(2), 553–609, 2018.
Piketty, T., L. Yang and G. Zucman, "Capital Accumulation, Private Property and Rising Inequality in China 1978–2015," *American Economic Review*, 109, 2469–96, 2019.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Appendix S1 Supporting Information