Working Paper Series

The impact of top incomes biases on the measurement of inequality in the United States

Vladimir Hlasny
Paolo Verme

ECINEQ WP 2017 - 452

ECINEQ 2017 - 452
November 2017

www.ecineq.org

The impact of top incomes biases on the measurement of inequality in the United States[*]

Vladimir Hlasny
Ewha Womans University, Seoul

Paolo Verme[†]
The World Bank, Washington DC, USA

Abstract

The paper re-estimates the Gini inequality index for the United States using the Current Population Survey between 1979 and 2014 and two alternative correction methods for top income biases. Top incomes in household surveys are affected by a variety of biases related to survey design, income measurement and post-survey data adjustments. The paper corrects for selected biases using a stochastic approach based on reweighting and a semi-parametric approach based on replacing observations. Both methodologies show that income inequality in the United States has been consistently underestimated during the period considered. The level of underestimation is up to 6.83 percentage points depending on the year considered, choice of correction method and choice of modalities within methods. The degree of underestimation is positively and significantly associated with mean income, non-response rates, the initial level of inequality and the degree of departure of the top incomes distribution from a Pareto distribution.

Keywords: Top incomes, inequality measures, survey nonresponse, Pareto distribution, parametric estimation.

JEL Classification: D31, D63, N35.

[*]The authors are grateful to Jessica Semega and Jonathan L. Rothbaum from the US Census Bureau for very helpful clarifications on the Bureau's methods, Jeff Larrimore from the Federal Reserve Board for helpful clarifications on topcoding, Ragui Assaad and Roy Van der Weide for peer reviewing an earlier draft of the paper, and to participants at the ECINEQ 2015 conference held in Luxembourg and the IARIW 2015 special conference held in Cairo for helpful discussions. All remaining errors are ours.

[†]**Contact details:** Paolo Verme, The World Bank, Email: pverme@worldbank.org.

1. Introduction

Top incomes have been in the limelight since the beginning of the global financial crisis in 2007 and the eruption of discontent that followed the crisis as expressed by the "We are the 99%" and "Occupy Wall Street" social movements. Inequality has for long been a central research theme in disciplines such as public economics, welfare economics or social choice theory and before the crisis several economists had started to focus on top incomes (Atkinson et al., 2011). However, it was the crisis that brought these works to public attention and saw a new push for research on top incomes and inequality. Joseph Stiglitz's book on inequality (2012) discusses at length the role of top incomes in rising inequality and how this fact compromises future prosperity in the US. Wealth inequality and accumulation at the top of the distribution is the central theme of Thomas Piketty's best seller "Capital in the 21st Century" (Piketty, 2014). In a 2013 article for the New York Times, Paul Krugman argued that "(…) *if you take a longer perspective, rising inequality becomes by far the most important single factor behind lagging middle-class incomes*"[3]. In a public speech delivered in December 2013, President Obama argued that income inequality "*(...) is the defining challenge of our time.*"[4] It is now publicly acknowledged that top incomes have grown disproportionally faster than other incomes during the past few decades, a phenomenon that seems common to developed and emerging countries alike.

Such phenomenon poses non-negligible problems to the measurement of income inequality. A few large incomes can significantly affect the measurement of income inequality (Cowell and Victoria-Feser 1996; Cowell and Flachaire 2007; Davidson and Flachaire 2007) and trends in incomes of the richest 1% of households have been shown to drive trends in income inequality over time (Burkhauser et al. 2012). The fact that top incomes are rising in numbers and weight and the fact that these incomes are difficult to capture in household surveys can potentially bias the estimation of income inequality significantly. Indeed, one of the important questions recently debated in various strands of the economic literature is how to correct survey data for top incomes biases.

National surveys suffer from a variety of issues related to the representation and measurement of top incomes (Groves and Couper 1998). These range from issues related to sampling (underrepresentation of the very rich) to issues related to data collection (unit nonresponse, item nonresponse, item underreporting

[3] New York Times (December 15th, 2013). http://www.nytimes.com/2013/12/16/opinion/krugman-why-inequality-matters.html?ref=paulkrugman&_r=0

[4] http://www.usatoday.com/story/news/politics/2013/12/04/obama-income-inequality-speech-center-for-american-progress/3867747/

and other measurement errors), data preparation (top coding, trimming or censoring, public provision of limited subsamples) or data analysis (trimming of outliers, choices of estimator).

Even the most sophisticated surveys such as the US Current Population Survey (CPS) – the official source for the measurement of income, inequality and poverty in the U.S. – suffers from various data issues such as under-reporting of government assistance programs (Tiehen, Jolliffe and Smeeding 2013; Meyer, Mok and Sullivan 2009; Meyer and Mittag 2014), top-coding of various components of income of high-income individuals (Burkhauser et al. 2011; Jenkins et al. 2011), and unit and item nonresponse particularly by high-income households (Korinek et al. 2006; Dixon 2007; Hlasny 2016).[5] The U.S. Census Bureau mitigates the prevalence of item nonresponses by imputing missing values using a "hot deck" method from responding individuals with similar socioeconomic characteristics. Nevertheless, this correction results in underestimation of inequality and overestimation of median income (Hokayem et al. 2016). Correcting for top income biases is therefore an essential exercise to measure the level and evolution of inequality over time. It is also an exercise that will become increasingly important if top incomes continue to grow disproportionally to the rest of the distribution.

The statistics and economics literature offer at least two classes of methods that try, with different means, to address the question of correcting inequality in the presence of top incomes biases. The first class relies on the comparison of between-survey and non-survey data. We can call this class of methods the *out of surveys correction methods*. Burkhauser et al. (2012), for example, compared tax-records and income-survey data and observed that they yield different measures of income inequality because of differences in income components and different definitions of inequality. Deaton (2005) showed how unit nonresponse may be one factor that can explain the discrepancy between national accounts and household surveys when it comes to the measurement of household consumption. A group of studies have attempted to align household survey results with those from national accounts by scaling up survey incomes to match aggregate national statistics (Bhalla 2002; Bourguignon and Morrisson 2002; Sala-i-Martin 2002).[6] Lakner and Milanovic (2013) combined corrections for unit nonresponse with corrections for measurement errors among top incomes and calibrated the estimated Pareto distribution among top incomes using aggregate income information from national accounts data.[7]

[5] See Annex for a detailed discussion of issues related to top incomes.
[6] This method avoids behavioral modeling of households' decisions, and hinges on the restrictive assumption that the difference between survey and tax-record incomes is distribution-neutral, so that it would be appropriate to scale up all survey incomes by the same factor.
[7] This method essentially assigns any disparity between the national accounts and household surveys to top-income households, effectively accounting for both unit nonresponse and measurement errors.

The second class of methods focuses instead on survey data only and tries to correct top income biases using within-sample information. We can call this class of methods the *within surveys correction methods*. There are two main methods under this class. The first method which we label *reweighting* aims at correcting the weights of existing observations using information on nonresponse rates across geographical areas. This method is used to correct for unit nonresponse (Mistiaen and Ravallion 2003; Korinek et al. 2006 and 2007) but can also be applied to item nonresponse. The second method, which we label *replacing*, aims at replacing top income observations with observations generated from theoretical distributions. This method is used to correct for issues such as top coding, trimming or censoring but can also be used for unit or item nonresponses if these nonresponses are concentrated among top incomes (Cowell and Victoria-Feser 2007; Jenkins et al. 2011).

This paper re-estimates the Gini inequality index for the United States using CPS data between 1979 and 2014[8] following the *within surveys correction methods* approach and comparing results using the *reweighting* and *replacing* methods. We find non-response rates to be positively and significantly related to income. We also find that both correction methods invariably increase estimates of the Gini coefficient for all years considered. With *reweighting*, the level of underestimation varies between 0.42 and 6.83 percentage points with a median value of 1.94 and a mean value of 2.34 across different waves of the CPS. *Replacing* based on the Pareto distribution provides much lower corrections with an upper bound of 0.39 percentage points.

The paper also tested and compared alternative flavors of the two correction methods considered and find a high degree of sensitivity to the different choices available. The wide range of results provided by the *reweighting* method stems from scale effects related to the size of incomes, the degree of non-response and the level of the uncorrected Gini and several differences across waves of the CPS including the geographic incidence of inequality and nonresponse across individual metropolitan statistical areas and states. *Reweighting* performs best in samples with low nonresponse rates, while it appears to be imprecise in samples with nonresponse rates over 10 percent. Analysis performed at an intermediate degree of geographic disaggregation (into states or clusters of 2-4 MCBSAs, of 388–1,697 households each) yields a better correction than those performed at hyper aggregation (Census divisions of 5,818 households each) or hyper disaggregation (individual MCBSAs, of 220 households each) degrees. The *replacing* method provides far lower corrections than the *reweighting* method. This would suggest that various top income

[8] Income data are collected in March every year and refer to incomes of the year before. Throughout the paper, we use the income year. For example, we refer to year 2014 when the survey is administered in March 2015.

biases operate in opposite directions. While unit nonresponse and possibly underreporting and top coding of incomes bias the Gini downward, the presence of extreme observations biases it upward.

The paper is organized as follows. Section two outlines the main methods used to correct for top income biases related to unit nonresponse. Section three describes the data. Section four provides the calibration of the models to the data at hand, section five presents the main results and section six concludes.

2. Models

This paper will cover and compare techniques that can be used to correct inequality estimates for top income issues related to unit nonresponse, item nonresponse, and measurement errors including top coding and trimming. As discussed, these techniques fall under two broad approaches: 1) *Reweighting* whereby original observations are kept intact while weights are recalibrated, and 2) *Replacing* whereby weights are kept intact but some observations are removed and replaced by others artificially generated. These two classes of techniques are presented below.

Reweighting

Reweighting is one possible approach to correct for unit nonresponse. Unlike the case of item nonresponse, we cannot simply infer households' unreported income from their other reported characteristics, because we don't observe any information for the non-responding households. Several statistical agencies have taken to the practice of assigning the mean or median values to the missing items, sometimes using the mean of the remaining observations in a cluster such as a Primary Sample Unit (PSU) and sometimes assigning the mean of the whole distribution. This is inappropriate, of course, as the missing values may be systematically very different from the rest of the cluster or distribution. In an effort to address this problem, Atkinson and Micklewright (1983) reviewed a method that relies on information about nonresponse rates across regions, whereas the mass of respondents in a region is 'grossed up' uniformly by the regional nonresponse rate. This is the approach essentially taken by the US Census Bureau in correcting the CPS (Census & BLS 2002, Ch.10-2). However, this approach is also problematic in that it accounts only for inter-regional differences in nonresponse rates, and not for systematic differences across units within individual regions.

Mistiaen and Ravallion (2003), and Korinek et al. (2006 and 2007) tried to address this last issue by using a probabilistic model that uses information on nonresponse rates across geographical units as well as information about the distribution within units.[9] It is assumed that the probability of a household i to respond to the survey, P_i, is a logistic function of its arguments:

$$P_i(x_i,\theta) = \frac{e^{g(x_i,\theta)}}{1+e^{g(x_i,\theta)}}, \qquad (1)$$

where $g(x_i,\theta)$ is a stable function of x_i, the observable demographic characteristics of responding households i that are used in estimations, and of θ, the corresponding vector of parameters from a compact parameter space. Variable-specific subscripts are omitted for conciseness. $g(x_i,\theta)$ is assumed to be twice continuously differentiable, but can take various functional forms. The parameters θ can be estimated by fitting the estimated and actual number of households in each region using the generalized method of moments (GMM) estimator

$$\hat{\theta} = \arg\min_{\theta} \sum_j \left[(\hat{m}_j - m_j) w_j^{-1} (\hat{m}_j - m_j) \right] \qquad (2)$$

where m_j is the number of households in region j according to sample design, $\hat{m}_j$ is the estimated number of households in the region, and w_j is a region-specific analytical weight proportional to m_j. The estimated number of households ($\hat{m}_j$) can be imputed as the sum of inverted estimated response probabilities of responding households in the region ($\hat{P}_{ij}$) where the summation is over all N_j responding households. If the sample is extracted from a larger population, the imputed true number of households should be divided by the sampling rate for the underlying population in each region (s_j) to obtain population estimates. Finally, if the available sample includes only a fraction of the households responding to the full survey in a region – such as a 25% random extraction from a sample – we should divide by the sample-extraction rate for each region (ss_j).

$$\hat{m}_j = s_j^{-1} ss_j^{-1} \sum_{i=1}^{N_j} \hat{P}_{ij}^{-1}. \qquad (3)$$

Under the assumptions of random sampling within and across regions, representativeness of the sample for the underlying population in each region, and stable functional form of $g(x_i,\theta)$ for all households and all regions, the estimator $\hat{\theta}$ is consistent for the true θ. Estimated values of $\hat{\theta}$ that are significantly different from zero would serve as an indication of a systematic relationship between household demographics and household response probability, and of a nonresponse bias in the observed distribution of the demographic

[9] Korinek et al. (2006, 2007) use the Current Population Survey, and correct for the presence of type-A unit nonresponse households. Mistiaen and Ravallion (2003) add the count of item nonresponse households to the type-A unit nonresponse households, and use the methodology on them jointly.

variable. In that case, we could reweight observations using the inverted estimated household response probabilities to correct for the bias.

Modalities of the method

Regional definition. The model presented in equations 1-3 above uses *within-j* information as well as *between-j* information. It uses within-*j* information because the estimated number of households $\widehat{m}_j$ is estimated within-*j* and it uses between-*j* information because the number of responding within-*j* households and the distribution of explanatory variables vary across *j*s. The choice of geographic disaggregation involves a trade-off between the number of *j* data points, and the number and distribution of within-*j* observations vis-à-vis the underlying population. On the one hand, observations should be behaviorally similar to non-responding households within-*j*, calling for smaller geographic units. The number of regions should also be sufficiently large, because model errors are at the level of regions *j*, and individual regions with atypical nonresponse rates or distributions of demographic variables should be prevented from exerting undue influence on estimates. On the other hand, equation 3 requires the sample to encompass the entire range of values of relevant characteristics of the underlying population, calling for larger geographic units.

Properties of the data at hand call for different degrees of data aggregation. Typical response rates, geographic variation in response rates, dispersion of incomes within and across regions, heterogeneity of households within regions, and the level of sample stratification are the parameters to consider. Korinek et al. (2006, 2007) used state-level disaggregation of CPS data, because geographic identifiers are consistently reported only at that level whereas county or metropolitan statistical area identifiers are missing for some responding as well as non-responding households. In their analysis of the Egyptian Household Income, Expenditure and Consumption Survey, Hlasny and Verme (2016) considered regional disaggregation both at a high administrative level (50 governorate by urban–rural areas with 939.7 observations on average) and at the level of primary sampling units (2,526 PSUs with 18.6 observations on average). These are clearly two different approaches with different implications. The PSUs tend to have relatively homogeneous within-*j* households, with similar behavioral responses between responding and non-responding households, and presumably similar survey-response probabilities. The observed range of household characteristics in each PSU is expected to comprise the values of non-responding households. A higher level of geographic aggregation would make behavioral responses less likely to be stable within areas *j*, while offering little additional assurance that values of characteristics of responding households encompass values of non-responding units.

Households' response probabilities are essentially inferred by comparing regions with similar ranges of explanatory variables. In the analysis at the finely disaggregated level, the response probability curve is constructed using numerous sets of probability estimates that are little overlapping on the curve. At the less disaggregated level, response probabilities are inferred by comparing fewer regions with greater ranges of incomes. The response probability curve is constructed using fewer sets of probability estimates largely overlapping. This paper considers alternative degrees of regional disaggregation to identify patterns in the correction for the unit nonresponse bias across the alternative specifications, and to identify the preferred degree of disaggregation for the data at hand.

Functional form. The relationship between households' characteristics and their response probability can be modeled in a number of ways including linear, logit or probit functions. This paper uses logit for its good empirical properties, for modeling convenience and in deference to previous literature. Furthermore, equation 1 allows various functional forms of household characteristics with varying degrees of allowed curvature, with or without monotonicity. Korinek et al. (2006, 2007) and Hlasny and Verme (2016) concluded that logarithmic specification of income yields better fit than linear, quadratic or higher-order polynomial forms, implying that unit nonresponse problem is concentrated in one end of the income distribution. This paper takes the problem of functional form in equation 1 as settled, and uses logarithmic form.

Explanatory variables. Korinek et al. (2006, 2007) evaluated a number of individuals' characteristics affecting their response probability, including income, gender, race, age, education, employment status, household size and an urban–rural indicator. Hlasny and Verme (2016) compared income and expenditures, and survey round. The choice over covariates involves a tradeoff between robustness of specifications with more covariates, and efficiency of more parsimonious specifications. Collinearity among covariates introduces estimation error. The above studies concluded that univariate models controlling for expenditures or income are the most efficient, prescription that this paper follows. The choice of income over expenditure or over income per capita may affect results systematically, since household-size adjusted incomes have been observed to be distributed more narrowly than household incomes.

Use of alternative household weights. The CPS offers a limited correction for unit nonresponse through sampling weights. Observations within adjustment cells (central and noncentral districts within metropolitan statistical areas, and urban and rural districts in non-MSAs) are reweighted by the density of nonresponding households in those cells (Census & BLS 2002, Ch.10-2). One issue is that this method accounts for differences in response rates across adjustment cells but not for systematic differences across income groups within individual cells. These sampling weights are problematic because they cannot be decomposed into weights for unit nonresponse and weights for other issues with unit representativeness, even when unit nonresponse is already addressed by the method discussed in this section. The options are either to essentially double-correct for unit nonresponse by using the available sampling weights, or to ignore other sample representativeness issues by not using the weights. This is an empirical question that depends on the magnitude of the respective issues with the sample at hand.

Korinek et al. (2007) considered both ways of weighting, and concluded that in the case of CPS the issue of unit nonresponse is more worrying than any issues with sample design and post-stratification, and so the unit nonresponse weights ($\hat{P}_i^{-1}$) should not be combined with the composite CPS weights. In what follows, two Gini estimates are produced: on uncorrected data, and on data corrected using statistical agency weights.

Replacing

Another body of literature argues that the best approach to correct for poorly reported or potentially contaminated top incomes is to remove the top end of the distribution altogether and replace it with synthetic values from a parametric distribution. Atkinson et al. (2011) summarize this literature contending that the distribution of top incomes is best approximated by a Pareto distribution (Pareto 1896) and use this distribution to model historical tax records in several countries. Cowell and Victoria-Feser (1996) and Cowell and Flachaire (2007) suggest combining parametric Pareto estimates for the top of the distribution with non-parametric statistics for the rest of the distribution. Testing this method on Egyptian data, Hlasny and Verme (2016) find that replacing actual top incomes with Pareto parametric estimates has a modest effect on the computed Gini. Hlasny and Intini (2015), using surveys from five Arab countries, concluded that the replacement with values from Pareto or generalized-beta distributions leads to different results across national surveys, but the alternative Gini estimates are within each other's confidence intervals. Burkhauser et al. (2010) compared four methods designed to address top-coding issues in survey data – essentially replacing top-coded values using four alternative parametric estimators – and combining the estimates with those from non-top-coded incomes. A more extreme approach has been recently proposed by Alvaredo and Piketty (2014) who ignore survey data altogether and propose to estimate inequality using a mix of Pareto distributions for top incomes and log-normal distributions for the rest of incomes. Using high parametric estimates for the distribution of top incomes as well as lower incomes, this approach yielded higher inequality measures for Egypt than those reported by Hlasny and Verme (2016). As discussed, we refer to these approaches as *replacing*.

We follow this literature to study the shape of the top income distribution and use the Pareto measures in two different contexts. First, we assess how sensitive Pareto coefficients are to unit nonresponse and other extreme income issues. Second, following Cowell and Flachaire (2007) and Davidson and Flachaire (2007) we correct the Gini coefficient for the potential influence of top observations by replacing potentially contaminated highest-income observations with values drawn from the expected distribution and combining the corresponding parametric inequality measure for these incomes with a non-parametric measure for lower incomes. The expected distribution of top incomes is obtained from a Pareto distribution right-censored at the income level above which income values are suspected of being contaminated. Finally, we compare the results with non-corrected Ginis or Ginis corrected for nonresponse using Korinek et al.'s (2007) method. This allows us to comment on the relative performance of the two alternative corrections for top-income biases.

The Pareto distribution is a particular type of distribution which is skewed and heavy-tailed. It has been used to model various types of phenomena and it is thought to be suitable to model incomes, particularly upper incomes. The Pareto distribution can be described by the following cumulative density function:

$$F(x) = 1 - \left(\frac{L}{x}\right)^{\alpha}, \qquad L \leq x \leq \infty, \tag{4}$$

where α is a fixed parameter called the Pareto coefficient and *x* is the variable of interest, in our case income, and L is the lowest value allowed for x The probability density function can be described as

$$f(x) = \frac{\alpha L^{\alpha}}{x^{\alpha+1}}, L \leq x \leq \infty \tag{5}$$

The probability density function has the properties of being decreasing, tending to zero as *x* tends to infinity and with a mode equal to the minimum value, L. As income becomes larger, the number of observations declines following a law dictated by the constant parameter α. Clearly, this distribution function does not suit perfectly all incomes under all income distributions, but it should be thought of as one alternative in modeling the right hand tail of a general income distribution.

If the income distribution is right-censored at H – separating *k* potentially contaminated top income observations from *n*-*k* reliable bottom observations – the probability density function becomes:

$$f(x) = \frac{\alpha L^{\alpha}}{x^{\alpha+1}} \Big/ \left(1 - \left(L/_{H}\right)^{\alpha}\right), L \leq x < H \tag{6}$$

where α in equation 6 can still be estimated using maximum likelihood from a right-truncated Pareto distribution.

Once we obtain α under the estimated right-truncated Pareto distribution, the Gini among the top k households ($H \leq x \leq \infty$) can be derived from the expression of the corresponding Lorenz curve (expression inside the integral below) as

$$Gini = 1 - 2\int_0^1 1 - [1 - F(x)]^{1-1/\alpha}\, dF(x) = \frac{1}{2\alpha - 1} \tag{7}$$

with a standard error composed of a sampling error in the estimation of the Pareto distribution, and an error in the estimation of the Gini coefficient. The sampling standard error under the Pareto distribution is equal to $4\alpha(\alpha - 1)/[\eta(\alpha - 2)(2\alpha - 1)^2(3\alpha - 2)]$ (Modarres and Gastwirth 2006), where η is the estimation sample size ($L \leq x < H$). The estimation error due to imprecision in the estimation of α is equal to $\epsilon/(2\alpha^2 - 2\alpha - 2\alpha\epsilon + \epsilon + 0.5)$, where ϵ is the standard error of $\hat{\alpha}$.

The parametric Gini coefficient from a Pareto distribution can be combined with the non-parametric Gini coefficient for the *n-k* lower incomes using geometric properties of the Lorenz curves to derive the semi-parametric Gini coefficient

$$Gini_{semi} = (1 + Gini_k)\frac{k}{n}s_k - (1 - Gini_{n-k})\left(1 - \frac{k}{n}\right)(1 - s_k) + \left(1 - \frac{2k}{n}\right). \tag{8}$$

Its variance is $\left[\varepsilon_k \frac{k}{n} s_k\right]^2 + \left[\varepsilon_{n-k}\left(1 - \frac{k}{n}\right)(1 - s_k)\right]^2$, where ε_k and ε_{n-k} are the standard errors of the two respective Gini coefficients, and s_k refers to the share of aggregate income held by the richest k households.

As long as it was correct to assume that top incomes in the population are distributed as Pareto, this semi-parametric Gini coefficient obtained with an estimated Pareto coefficient α can be compared to an uncorrected non-parametric estimate for the observed income distribution. A difference between the semi-parametric and non-parametric estimates would indicate that some observed high incomes may have been generated by a statistical process other than Pareto, or that some upper income values are missing due to nonresponse, and that the inequality index is sensitive to these issues. Semi-parametric Gini that is lower than the non-parametric Gini can be interpreted as evidence that some top incomes in the sample are ‘extreme’ compared to those predicted under the Pareto distribution. Higher semi-parametric Gini would indicate that the observed top incomes are distributed more narrowly than what the Pareto distribution would predict, potentially implying under-representation, censoring, or measurement errors among high-income units in the sample.

Modalities of the method

Estimation of parameters. In this paper, we estimate α in equation 6 directly using maximum likelihood from a Pareto distribution right-censored at an income level above which income values are suspect, such as the 95th percentile of income. We obtain an estimate with a robust standard error (Jenkins and Van Kerm 2007). Another possible definition of the Pareto coefficient (α) as well as the inverted Pareto coefficient (β) as proposed in Atkinson et al. (2011) would be:

$$\alpha = \frac{1}{1 - \left[\log(\frac{s10}{s1}) / \log(10)\right]} \tag{9}$$

$$\beta = \frac{\alpha}{\alpha - 1} \quad or \quad \alpha = \frac{\beta}{\beta - 1}, \tag{10}$$

where $s10$ and $s1$ represent the income shares of the top, say, 10% and 1% of the population respectively. With tax records, it is generally more common to use the top 1% and 0.1% respectively but with household data, where samples are typically in the thousands of observations, the top 0.1% of households is a sample too small to be representative of the very top of the distribution as it may comprise extreme observations, hence the choice of the top 1% of the population, assuming that these top 1% of observations are more representative of the true top incomes.

The interpretation of the beta coefficient is that larger betas correspond to larger top income shares while the opposite is true for the alpha coefficient. As a rule of thumb, the beta coefficient is what provides a snapshot indication of top incomes. Research on top incomes has shown that the alpha and beta coefficients are rather stable across income distributions, in any given year and country, as originally predicted by Pareto. The work by Piketty and others, which used much longer time-spans than previous research, has shown that the beta coefficient can vary over time and that this variation can be explained by a combination of economic and political factors.

Cowell and Flachaire (2007) propose the following formulation of α

$$\alpha = \frac{1}{k^{-1} \sum_{i=0}^{k-1} \log X_{(n-i)} - \log X_{(n-k+1)}}, \tag{11}$$

where $X_{(j)}$ is the jth order statistic in the sample of incomes n, and k is the delineation of top incomes such as the top 10% of observations.

Alternative parametric modeling. While the Pareto distribution is thought to approximate well the dispersion of top incomes, it is not representative of incomes in the middle or bottom of the income distribution. Generalized Beta distribution of the second kind (GB2), also known as the Feller-Pareto distribution, has been proposed as a suitable functional form representing well the vast bulk of the income distribution (Thurow 1970; McDonald 1984). With appropriate parameters, the upper tail of the distribution can be heavy and can decay like a power function. The lower end of the distribution can be short-tailed. Four estimable parameters give the distribution flexibility to fit various empirical income distributions. The cumulative distribution function of the GB2 distribution is

$$F(x) = I\left(p, q, \frac{(x/b)^a}{1 + (x/b)^a}\right) \tag{12}$$

where *I(p,q,y)* is the regularized incomplete beta function, in which the last argument, *y*, is income normalized to be in the unit interval. Parameters a, p, and q are distributional shape parameters and b a scale parameter that can be estimated by maximum likelihood. Other suitable candidates for a distribution function, the Singh-Maddala (1976) and the Dagum (1980) distributions, are limiting cases of the GB2 distribution with parameter p (q, respectively) restricted to unity (McDonald 1984).

Because bottom incomes may not be approximated well by the GB2 distribution, Jenkins et al. (2011:69) proposed estimating the GB2 distribution on income data left-truncated at the 30th percentile. Similarly, because the right tail may be contaminated by top income issues, right-truncation may be applied in the calculation of the GB2 density and model likelihood functions (see Annex table A2, and footnote 26). The Gini index of income inequality under the estimated (truncated) GB2 distribution can be computed by evaluating the generalized hypergeometric function ${}_3F_2$ with the estimated parameters as arguments. Its standard error can be computed using the delta method (McDonald 1984).

One issue with replacing of potentially imprecise true top incomes with fixed Pareto fitted values is that the resulting measures of income distribution and inequality do not account for parameter-estimation error and sampling error in the available sample. An and Little (2007), and Jenkins et al. (2011) account for sampling error by drawing random values from the estimated distribution for all potentially imprecise top incomes, calculating a quasi-nonparametric inequality measure with its standard error, repeating the exercise multiple times and observing variability in the obtained inequality measure.[10] Following Reiter (2003), the expected measure of inequality in such 'partially synthetic' data can be computed as a simple mean of inequality measures from individual random draws:

$$\widehat{Gini_q} = \sum_{i=1}^{m} Gini_{qi} / m \tag{13}$$

In this expression, $Gini_{qi}$ is the quasi-nonparametric Gini coefficient from a random draw i, and m is the number of draws. Sampling variance of the expected $\widehat{Gini_q}$ index can be computed as:

$$\widehat{var} = \frac{\sum_{i=1}^{m} \left(Gini_{q\,i} - \widehat{Gini_q}\right)^2 / (m-1)}{m} + \sum_{i=1}^{m} var_{qi} / m. \tag{14}$$

[10] Since top incomes in the US-CPS do not appear to follow Pareto distribution exactly, Jenkins et al. fit the GB2 distribution instead. Since top-coding occurs at the level of individual components of income, this estimation is done at the level of income components, and the randomly drawn values for top coded components are added to actual values for non-top coded components.

The first term is the sampling variance across different draws from the Pareto distribution, and the second term is the mean sampling variance within an individual draw. *m* refers to the number of repetitions, and var_{qi} is the variance of the quasi-nonparametric Gini coefficient from an individual draw *i*. This methodology still ignores standard error from the estimation of parameters in the Pareto distribution. However, this standard error is expected to be quite small compared to the sampling error, and can be ignored in large datasets where parameters have been estimated precisely (Jenkins et al. 2011).

Selection of a threshold for estimation, and for replacement of top incomes. The Pareto distribution has been found to represent well the dispersion of topmost incomes, but its fit becomes worse for incomes closer to the mean. In this study, we estimate the parameters of the Pareto distribution among the top 15–25 percent of incomes.

While lower incomes provide a poor fit to the Pareto distribution because of the distribution's simple parametric shape, the highest incomes also present a problem due to their potential contamination with top income biases. Upper truncation is warranted at a level above which income observations are feared not to represent the true population incomes. Cowell and Flachaire (2007) propose a threshold at the 90th percentile of incomes. Jenkins at al. (2011) raise a concern that top income issues are pervasive across a wide range of US incomes, affecting some income categories among the two upper deciles and perhaps even the top one half of incomes. On the other hand, on the basis of the quality of fit in the United Kingdom income surveys, Jenkins (2016) advocates setting the threshold at 1% or 5% of top incomes.

We replace between the top 1% and 10% of incomes with uncontaminated synthetic values. Our motives are empirical. Even at this level of wealth, the US population and the CPS sample have a sizable number of individuals, and a large fraction of these are regular workers who respond to income surveys and do so truthfully. The top 1 percent represents 3.23 million people in the US. In the 2012 income data from the CPS,[11] the top 1 percent of incomes includes 752 surveyed households earning over $356,159. The following 4 percent includes an additional 3,093 households earning over $192,144, and between the 90th and 95th percentile, there are 3,886 additional households earning just $147,001 or more. Moreover, the threshold for replacing values should be set relatively high so that as large a number of income observations as possible could be used to estimate the Pareto distribution precisely. Setting the threshold for replacing values any lower than the 90th percentile could affect the degree of fit adversely.

[11] In the following analysis, all years refer to income reference years, for which data come from the CPS March Supplement from the following year.

3. Data

The paper uses income data from the 1979–2014 Current Population Surveys, March Annual Social and Economic Supplement produced by the United States Census Bureau. We use the data files described as “public access” data, available on-line from the US Census Bureau’s or the National Bureau of Economic Research’s (NBER) websites.[12] They should be distinguished from the “restricted access” data, which are provided by the US Census Bureau to selected researchers in secure offline locations, following an assessment of their research proposals. All data sets and variables used in this paper are fully documented with ancillary files available on-line from the US Census Bureau’s and NBER’s websites.

As a measure of wellbeing, the paper uses household income not adjusted for either household size or composition.[13] This is the main indicator used by the US Census Bureau to measure the Gini index (US Census Bureau, 2000). The Bureau also publishes an alternative series based on household equivalent scales.[14] Both choices have their own drawbacks.[15] However, as our objective is to re-estimate official US figures, we will adopt the main indicator used by the Bureau.

The public and restricted access files are both topcoded, which means that researchers have no access to non topcoded files. The difference between the two versions of the data is that topcoding is heavier in the public access files (Figure 1). Topcoding affects all years between 1979 and 2014, although topcoding methods have changed over time. The latest technique used for top-coding starting from 2010 is “rank proximity swapping,” whereby values above the cutoff for top-coding are swapped one for another within bounded intervals. As a result, the imputed values are similar but not identical to the latent true values.[16]

[12] Available at: https://thedataweb.rm.census.gov/ftp/cps_ftp.html and http://www.nber.org/data/current-population-survey-data.html.

[13] The whole analysis was also replicated on household income per capita, and family income. For a justification of the household income per capita, see among others the life works of Anthony Atkinson, Francois Bourguignon, Martin Ravallion, Branko Milanovic and Joseph Stiglitz.

[14] See detailed explanations here:
https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/equivalence.html

[15] Using household income as opposed to household income adjusted for household size (per capita) or composition (per adult equivalent) carries the implicit assumption that income is spent entirely on public goods (each household member disposes equally of the full household income). This is of course not true. Some goods such as housing and utilities are clearly a public good but others such as personal services, clothing, food or mobile phones are not. It also assumes that each household member has the same consumption capacity, which is also not true. Adults eat more than small children, and men and women have different consumption patterns. Adjusting for household composition has also its own drawbacks. Consumption patterns and family structures change across time, regions and countries. If one is really concerned about comparing household wellbeing across time and space, the scales used to account for household composition should be adjusted across these two dimensions. This, of course, becomes very cumbersome and increases the degree of normative choices that the analyst is called to make. On the other hand, using one equivalence scale for all would be very controversial and would favor some countries at some point in time at the expenses of other countries or points in time (different organizations such as the OECD or the FAO have proposed different scales but there is no international agreement on the most appropriate scale to use).

[16] As described in US Census Bureau (2015) “The 2015 ASEC public use data file uses a method that swaps values between sample cases having incomes above a determined topcode value. This method of topcoding preserves the distribution of values above the topcode while maintaining adequate disclosure avoidance. The technique used for swapping values is termed “rank proximity

Total household incomes and incomes per capita imputed from them could differ from the true values for a substantial fraction of the sampled households (Jenkins et al. 2011).[17] Indeed, using the public data files, we have estimated the total number of households in the nation and reproduced exactly the number of households published by the Census Bureau. However, when we estimate the number of households by income group, also published by the Census Bureau, our estimations are exact only for very low income groups. This is the effect of top-coding and has an obvious impact on the estimate of inequality.

We find top-coding to systematically reduce the Gini coefficient. Figure 1 compares the Gini index estimated with the public files with the one published by the Census Bureau, both estimated on household income with household weights. Figure 1 also shows the extent of topcoding of public-access versus restricted-access Census Bureau data (Larrimore et al. 2008). Our Gini estimates (Gini, public access data) follow very closely the trend of the published Gini and the difference between the two series is rather constant, at least for income reference years 2000 to 2014. During this period, the difference between the two Ginis varies between 0.7 and 0.9 percentage points.[18] This is to some extent the effect of top-coding on the Gini. Therefore, our top income corrections are applied to an initial Gini that is underestimated because of topcoding. Note that, in 1985, topcoding was of the same magnitude between the public and restricted-access data, and they were similar during the period 1982–1986. This indicates that our correction for top income biases in 1985 is the same as if we were using the restricted access data. It is also very similar for the period 1982-1986.

In the following cross-section analyses, the 2012 CPS March Supplement survey is used as benchmark study. This choice was made for a variety of reasons. The survey is relatively recent and follows the latest change in the CPS topcoding method introduced in 2010. It is also representative of the entire series. The estimated unit non-response bias for 2012 is very close to the mean value estimated for the period 1979–2014 and the US nominal GDP growth for the period 1979–2014 is very close to the one observed for 2012 and 2013. Moreover, the year 2013 saw a change in the survey methodology that led administrators to

swapping". Once the topcode has been established, individuals with value above the topcode cutoff are sorted by those values from lowest to highest (values equal to the specified topcode are included in the universe of those requiring topcoding). Next, the values above the topcode are systematically swapped between sample persons. The swapping occurs within a bounded interval. This bounded interval assures that the values swapped are in 'proximity' to each other, yet providing a sufficiently large group of persons from which the swap partners are selected." Beside swapping, the imputed values are rounded to two significant digits (e.g., $987,654=$990,000; $12,345=$12,000; $9,870=$9,900 – Refer to CPS 2013, Chart 1).

[17] For more information on the US-CPS see https://www.census.gov.

[18] It should be noted that there are two relevant breaks in series that affected top-coding. In 1993, the Census Bureau increased the top-coding limit on many income items (see footnote 23 at https://www.census.gov/topics/income-poverty/income/guidance/cps-historic-footnotes.html) and for year 2013 it redesigned the survey to better capture retirement, interest, and dividend income (footnotes 38 and 39). For details on top-coding practices through the years, see also: Historical CPS ASEC Income Extract Files using Topcode Proximity Swapping 1975-2010 (on the page https://www.census.gov/data/datasets/time-series/demo/income-poverty/data-extracts.html).

provide two versions of survey data, one traditional and one redesigned, each on a limited sample size. This made the latest rounds of the CPS more complex to use for cross-section analyses.

Table A1 in annex reports the number of responding households, nonresponse rates, mean income and the Gini index by state for the 2012 round. Nonresponse rates vary between 4.1 percent in Montana to 15.4 percent in Maryland with a national value of 9.5 percent.[19] Incomes exhibit a high degree of inequality across all states with a mean state value of 45.42 and clear differences across states (standard deviation 2.69). The lowest level of inequality was recorded in Nebraska with a Gini of 41.20 and the highest in New Mexico with a value of 54.75. A decomposition of inequality between and within states shows that the pure within-state component accounts for only 1.4 percentage points of the Gini and the pure between-state component accounts for 7.9 percentage points, while their overlap accounts for 37.0 percentage points.[20]

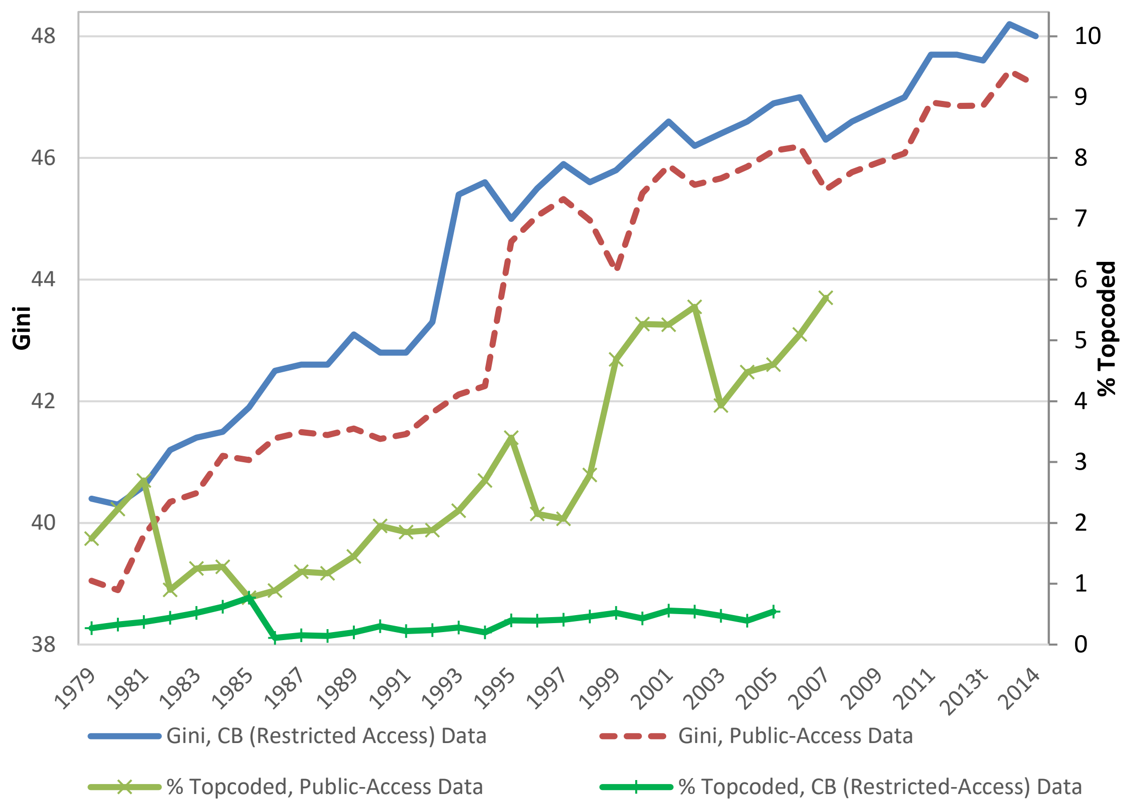
We provide all estimates with and without data weighted by the sampling weights provided by the US Census Bureau. These weights are the product of several adjustments including a partial adjustment for nonresponses. Given that we are correcting for nonresponse rates, using these weights would lead to double correction. To estimate the size of this double correction we provide a simple Difference in Difference test and we also report all results with and without sampling weights.[21]

Figure 1 – Gini Index Estimates and Percent of Individuals Topcoded (income years 1979–2014)

[19] The US Census Bureau distinguishes three types of unit non-interviews: explicit refusals or absence of anyone at home (type A), and vacant, demolished or otherwise un-contactable units (types B and C). Here we restrict our attention to type A nonresponse following Korinek et al. (2007). Nonresponse rates are theoretically available at any geographic level, because CPS reports geographic indicators even for nonresponding units. However, more detailed geographic indicators – those for counties or metropolitan areas – are sometimes missing for nonresponding as well as responding units, and only state is available for all units.

[20] The results on unweighted data, or data weighted by nonresponse correction weights are analogous.

[21] According to the US Census Bureau "*The final weight, which is the product of several adjustments, is used to produce population estimates for the various items covered in the regular monthly CPS. This weight is constructed from the basic weight for each person, which represents the probability of selection for the survey. The basic weight is adjusted for special sampling situations and failure to obtain interviews from eligible households (noninterview adjustment). A two-stage ratio estimation procedure adjusts the sample population to the known distribution of the entire population. This two-stage ratio estimation process produces factors which are applied to the basic weight (after the special weighting and noninterview adjustments are made) and results in the final weight associated with each record. In summary, the final weight is the product of: (1) the basic weight, (2) adjustments for special weighting, (3) noninterview adjustment, (4) first stage ratio adjustment factor, and (5) second stage ratio adjustment factor. This final weight should be used when producing estimates from the basic CPS data*" (US Census Bureau 2015). For unit nonresponses, the Bureau reweights observations within adjustment cells (central and noncentral districts within metropolitan statistical areas, and urban and rural districts in non-MSAs) by the density of non-responding households. This accounts for differences in response rates across adjustment cells but not for systematic differences across income groups within individual cells. Korinek et al. (2006) find this limited correction method to be inadequate for top income distribution, while Bee et al. (2015), by linking CPS data to Internal Revenue Service Form 1040 records using individuals' addresses, allege that the Bureau's method yields income distribution with correct quantiles. One limitation of the latter study is that the proportion of linked households falls dramatically among the richest and poorest 5 percent of households. Bollinger et al. (2017), linking CPS data to Social Security Detailed Earnings Records, find that unit nonresponse in the bottom and top tail of the income distribution is responsible for one-half of the differences in inequality estimates between the two data sources. Poor income measurement is responsible for the remaining differences in inequality measurement across data sources. Juster and Kuester (1991) find that different household income surveys provide significantly different estimates of the income distribution due to different degrees of misreporting of various income components, unit and item nonresponse, and sample attrition rates.



Source: Authors' estimates (Public Data) and US Census Bureau (CB Data), Table H4 available at: https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-households.html. Income reference years are shown (with values taken from the following years' March CPS). Percentages topcoded are adopted from Larrimore et al. (2008).

4. Models calibration

Reweighting

The proposed reweighting model requires several important decisions that affect results including the choice of functional form, explanatory variables, weights, and geographic disaggregation level. The former three choices have been resolved in previous contributions and are discussed in the models section. The choice of optimal disaggregation level, instead, has not been treated in previous literature and needs to be addressed here. The number of regions *j* selected for the estimation of equation 2 determines the weight that the model attributes to *within*-region as opposed to *between*-regions information and this choice leads to significantly different estimations of the correction bias. Therefore, before applying the reweighting method to the US CPS data series, we need to test i) how the bias correction changes as we change the disaggregation level; ii) what is the optimal level of disaggregation.

To test how the bias changes as we change the regional aggregation level, we use a subsample of the 2012 CPS that allows to construct a sufficiently large numbers of regional stratifications each with a nonresponse rate.[22] Table 1 reports the results. The table shows that the more detailed the degree of geographic disaggregation, the smaller the estimated bias due to unit nonresponse (with few exceptions, such as the analysis performed at the level of Census divisions, column 2). In the data uncorrected using sampling weights, the bias estimated using state-level disaggregation is 2.91 percentage points, falling to 0.99 points when estimated using MCBSA-level disaggregation. This is not due to any systematic selection of households between those reporting and those non-reporting their MCBSA. The samples used in columns 2 and 3 are identical. Also, adding together households reporting and households non-reporting their MCBSA (24 states analysis – column 1), we obtain similar results as when we restrict our attention to households reporting their MCBSA (column 3). This suggests that the availability of information on households' MCBSA is not systematically related to their income or their tendency to respond to the survey.

[Table 1]

Analyses using finer degrees of disaggregation typically yield lower corrections for unit nonresponse for several reasons. One, finer degrees of disaggregation translate into more numerous and smaller error terms in equation 2. This prevents any group of regions with outlying values of nonresponse rates or extreme incomes from unduly influencing the estimable relationship in equation 1, and allows more precise estimation of all statistics. Indeed, coefficient standard errors are significantly lower when finer degrees of disaggregation are used. Two, finer disaggregation reduces the dispersion of incomes within regions and reduces the overlap of income distributions across regions, particularly in datasets where inequality abounds at a lower geographic level rather than across different parts of the country. This reduction in dispersion within regions and in overlap across regions restricts the mechanism in the task of reweighting observations (equation 3), because greater fractions of observations in each region must be assigned similar weights, including very high or very low weights under the common response-probability function estimated for all regions. This is particularly restrictive in datasets with little overlap in income ranges and modest differences in nonresponse rates across regions.

[22] The sample includes 24 states, each with MCBSA and nonresponse information available for over 75% of sampled households. The CPS also includes information on counties, but for fewer than 50% of households, and so this information cannot be effectively used.

Indeed, the change in the estimated bias across different disaggregation levels is notably large in the US, where substantial income inequality exists at the sub-state level, across cities rather than across states, and where nonresponse rates are similar across states as well as across MCBSAs. As mentioned above, disaggregation from the state to the MCBSA level (eight times smaller regions) reduces the estimated bias to the Gini coefficient from 2.91 to 0.99 percentage points, by 66 percent. At the level of MCBSAs, nonresponse rates range from 0.0 to 23.5%, compared to 4.1–15.4% at the state level. Since the degree of geographic disaggregation of survey sample affects the correction for unit nonresponse systematically, the natural question then arises as to what geographic disaggregation would produce the most appropriate correction.

The model in equations 1–3 relied on two assumptions about the underlying population and the sample: stability of the behavioral response across responding and non-responding households as well as across regions; and representative sampling across all income strata in the population. These conditions prescribe what the composition and the disaggregation of the sample should be. On the one hand, observations should be behaviorally similar to non-responding households within-j, and to households with similar values of income in surrounding areas, calling for smaller geographic areas. For the imputation of response probabilities, it is more meaningful to compare the frequencies of observing incomes of households with their counterparts in neighboring areas within a part of the country, than with households from across different parts of the country. On the other hand, equation 3 requires that the sample of respondents be representative of all population strata and encompass the entire range of incomes of non-respondents, potentially calling for larger geographic areas. Geographic regions should thus be small but not too small.

To test for the optimal level of disaggregation, we conduct a simple experiment. We first choose a high quality CPS sample with low nonresponse rates and low estimated biases due to nonresponse. We then trim observations across the distribution using a response-probability function based on income so that higher income households are more likely to be excluded. We treat these trimmed observations as the only source of unit nonresponse to correct for, essentially using the original observed sample and Gini as untainted by unit nonresponse. Finally, we use the reweighting procedure illustrated in section 2 to correct the Gini and compare this Gini with the one based on the full sample.

For this exercise, we again use the 2012 sample of the CPS. 73.5 percent of responding households and 79.2 percent of non-responding households have a known MCBSA. We can evaluate the correction for unit nonresponse bias performed at the level of Census regions, states and groups of MCBSAs.[23]

[23] The same analysis was performed on twenty US states, each with over 80% of households with known MCBSAs, with essentially the same results as in table 3. The analysis was also performed on the 2009 income reference year, covering one of the largest

For the trimming of observations, we apply the stochastic behavioral response proposed in equation 1 and validated in table 1. Richer households have a lower propensity to appear in the sample. Households' probability of response – and thus one minus the probability of being trimmed – is made a logistic function, with a simple logarithmic function of income in the numerator and the denominator. Using coefficient estimates in table 1 and similar model specifications considered (refer to tables 4 and 5), $g(x)$ in equation 1 is taken to be: $g(income)=\theta_1+\theta_2 log(income)=15.0-1.1 log(income)$.

Table 2 reports the estimated Ginis after reweighting with their standard errors. The difference between the Ginis biased due to trimming and the corrected Ginis is also shown. A difference of zero represents perfect correction, while a positive value represents undercorrection and a negative value represents overcorrection. The exercise is repeated across different degrees of geographic disaggregation (columns 1–6) for different percentages of observations trimmed (from 5% to 16%, row panels).[24]

[Table 2]

The reweighting method proposed invariably reduces the top income bias. The difference between the true Gini and the corrected Ginis is always significantly smaller in absolute value than the bias induced by trimming. The correction appears to perform better when the model is run at an intermediate degree of geographic aggregation. Models run at extreme aggregation levels (7 census divisions or 185 MCBSAs) never produce the best corrections, and always under-correct for the bias. The best corrections are found in between the models run at the state level (24 states) and those run at the level of groups of 2–4 MCBSAs (61–103 regions). Either of these intermediate disaggregation levels can be justified as performing well under some trimming scenario. However, as the percentage of trimmed observations increases, the levels of aggregation that perform best become systematically concentrated around the center of the spectrum of possible aggregation levels (41–103 groups of 2–10 MCBSAs). These findings hold whether sampling weights are used or not. The implication is that, with arbitrary empirical incidence of household nonresponse, particularly when the aggregate nonresponse rate is high, corrections that try to fit the population in a sufficient number of regions (i.e., 24 or more here) by reweighting a sufficient number of income observations per region (i.e., 388 or more), appear to yield the best corrections.

While the analyses performed at the levels of Census divisions, states or individual MCBSAs yield somewhat inconsistent corrections, the analyses performed at the levels of groups of MCBSAs exhibit high

samples across years, with one of the lowest nonresponse rates in recent years and one of the lowest estimated Gini biases due to unit nonresponse. This sample is as close to one free of unit-nonresponse problems as we could find. The results are estimated only slightly more precisely than those for year 2012.

[24] The algorithm performing randomized trimming according to household weights could not trim a smaller number of observations while observing the desired weighting scheme.

consistency in relation to one another. In this range of geographic disaggregation, the more detailed the disaggregation, the lower the correction of the bias, corroborating the evidence in table 1. Regarding derivation of the actual behavioral response function, the models fitted in table 2 perform well when 5–7% of the sample is trimmed, in models using disaggregation into a sufficient number of regions. Estimated coefficients are within one standard deviation from the actual values (θ_1=15.0, θ_2=-1.1). When more observations are trimmed, or when Census-division disaggregation is used, estimates differ from the actual values more, suggesting poor fit.

Finally, both tables 1–2 indicate that the unit-nonresponse bias corrections are higher when sampling weights are applied in the sample. CPS sampling weights provide limited correction for non-representativeness of sampled units, including unit nonresponse, stratified sampling and other issues with the sampling frame, particularly in relation to top income individuals. These weights tend to assign higher mass to rare rich individuals, so when they are multiplied by the nonresponse correction weights, they yield even higher bias corrections. This is expected, since various top income biases operate simultaneously and even in a complementary fashion among high income observations.

In conclusion, table 2 provides several important insights regarding the performance of the method for correcting for the unit nonresponse bias through reweighting of income observations. The method performs best in samples with low nonresponse rates, while it appears to be imprecise in samples with nonresponse rates over 10 percent. Analysis performed at an intermediate degree of geographic disaggregation (into 24–105 regions, states or clusters of 2–10 MCBSAs, of 388–1,697 households each) yields a better correction than ones performed at hyper aggregation (Census divisions) or hyper disaggregation (individual MCBSAs) degrees. Overall, properties of the data at hand should guide the choice over the appropriate degree of disaggregation and proper weighting of data.

Replacing

Similarly to the reweighting model, the proposed replacing model also requires several important decisions including the choice of parametric form, weights, points of truncation for the estimation of the parametric distribution, and the cutoff for incomes to be deemed uncontaminated by top income issues. This section explores the sensitivity of the results to these modalities. Regarding parametric form, the one-parameter Pareto distribution (type 1) and the four-parameter generalized beta distribution (type 2) are evaluated. Alternative left truncation points for the distributions are considered, as are right truncation points which also serve as the cutoffs for contaminated incomes to be replaced by values drawn from the parametric distributions. Finally, alternative data weighting schemes are applied.

Table 3 presents semi-parametric estimates of Gini coefficients obtained by replacing the highest top 1, 5 or 10 percent of potentially contaminated income observations with values imputed from a Pareto distribution as per Cowell and Flachaire (2007), and Davidson and Flachaire (2007).[25] The three blocks of rows present the results of the estimation of the Pareto distribution among the highest 15, 20 or 25 percent of incomes, respectively, to evaluate the fit of the Pareto distribution on differently delineated groups of high incomes. The samples used for estimation are right-truncated at the 99th, 95th or 90th percentile – refer to individual rows in table 3 – to explore the degree of potential contamination of the topmost incomes. Finally, actual observations of the bottom 99, 95 or 90 percent of incomes are combined with random draws from the Pareto distribution for the top 1, 5 or 10 percent of incomes, respectively, to obtain semi-parametric estimates of the Gini.

[Table 3]

Overall, the bias is never negative confirming that the US-CPS estimated Gini is underreported. The bias is also clearly smaller than the one obtained from reweighting. Across all rows of table 3, the estimated Pareto coefficient α ranges from 1.103 to 2.403, and the corresponding inverted Pareto coefficient β ranges from 1.713 to 10.696 depending on whether only the top 15% or up to the top 25% of households are used for estimation. The wider the domain of incomes used, the typically higher the estimated dispersion among top incomes and the higher the estimated top-income share. This could indicate that incomes around the 75th and 80th percentile are particularly widely dispersed, or that they are not represented well by the Pareto distribution, in general agreement with the claim that the Pareto distribution provides especially good fit at the top of the income distribution, but not necessarily below. Among the evaluated choices, for conceptual and empirical reasons, the most appropriate value for left truncation of the Pareto distribution on the CPS sample thus appears to be the 85th percentile. A lower left-truncation point makes Pareto estimates susceptible to mis-estimation due to the poorer fit of the parametric distribution to lower income observations. A higher left-truncation point leads to a reduction of sample size, particularly given that right-truncation is simultaneously applied at a level where incomes are suspected of being extreme or tainted by measurement issues (say, the 95th percentile).

[25] Beside these cutoffs, we considered replacing the highest top 0.1–20% of income observations with values under the Pareto distribution. These various percentages of top incomes to be replaced are in recognition that extreme observations of various income components – and top-coding of these observations in US-CPS – occur even among households with total incomes that do not appear extreme (Burkhauser et al. 2011). The effects of replacing only 0.1-0.5 percent of income observations are negligible compared to those in table 3, as they reflect the replacement of very few observations. On the other hand, replacing of up to 20% of top incomes yields unexpectedly low Pareto coefficients (0.66-1.25) and negative or extremely high inverted Pareto coefficients, confirming that the fit of the Pareto distribution deteriorates when the domain of incomes under consideration widens. The results are in line with those presented in table 3 and are thus omitted.

Another finding is that the more of the highest incomes are replaced with draws from the Pareto distribution, the higher the estimated Gini. The actual replaced values are distributed more narrowly than predicted under the Pareto distribution. Income underreporting or nonresponse among households earning the top 1–10 percent of incomes may be responsible. Another contributing factor may be that topmost incomes in the CPS are top-coded – although the rank-proximity swapping and rounding method does not systematically lower the treated income values. Since the Pareto distribution provides the best fit among the highest income observations, these findings indicate that a sufficient number of high income observations should be retained in the sample to estimate α that will facilitate good prediction to out-of-sample, uppermost observations. This suggests a right-truncation point and a cutoff for replacement at the 95th or a higher percentile of incomes.

Finally, table 3 indicates that the estimated Ginis depend on the way income observations are weighted. Ginis estimated on a sample weighted by CPS sampling weights are systematically higher than Ginis on an unweighted sample. Moreover, as in tables 1–2, the estimated top-income bias corrections to the uncorrected Ginis are higher in the weighted sample. Because the CPS sampling weights tend to assign higher mass to rare rich individuals, their use typically results in an increased estimate of inequality. At the same time, because our method corrects the values of high incomes, it operates in a complementary fashion with this increase in the mass of rich individuals, and the correction is higher on weighted data.[26,27]

To summarize, the corrections for potentially extreme or imprecise top income observations through replacing – of up to +1.93 percentage points under Pareto distribution – are lower than the corrections for unit nonresponse through reweighting. Parameter estimates are also sensitive to the sample used for estimation, that is, the left and right truncation points. Regarding the most appropriate specification of the replacing model using Pareto distribution, results in table 3 suggest that Pareto coefficients should be estimated only among the rightmost values of incomes that are clear of contamination by top income biases,

[26] Sampling weights are likely to be higher among individuals whose incomes were top-coded out of privacy concerns, since these individuals are rare. Our method tends to replace their income with parametric values, and at the same time retains their sampling weights.

[27] Replacement of top incomes was also conducted using the left- and right-truncated generalized beta (type 2) distribution (Table A2 presents basic results). Contrary to the Pareto replacement, GB2 replacement yields Ginis lower than their uncorrected values. Actual top incomes appear to be distributed more widely than predicted under the GB2 distribution. Since we suspect top incomes to suffer from a variety of biases, including a downward nonresponse bias and potentially an upward bias due to extreme values, the GB2 estimates would suggest that the upward biases are substantially greater than any downward biases, a dubious proposition. The parameter estimates are sensitive to the left and right cutoff of the estimation sample. This suggests that the four estimated GB2 parameters adjust to provide good in-sample fit at the expense of the out-of-sample prediction to the rightmost tail. For these reasons, we reject the GB2 distribution as providing acceptable fit for the topmost incomes. The estimates of the Gini under the GB2 distribution are too low and are volatile under alternative model specifications. On the other hand, the Pareto distribution is viewed as giving more realistic and stable predictions for the uppermost tail of the income distribution, supporting the conjecture that the parsimonious Pareto distribution provides a superior fit at the right tail, over the GB2 distribution that may fit in-sample lower-income data better but produces poor out-of-sample predictions in the rightmost tail.

such as those in the 85th to the 95th percentile. To the extent that replacing generally corrects for different sources of biases than sampling weights, the method should be applied on weighted data.

5. Results

Given the findings in the above calibration exercises, this section presents the results of model specifications considered as the most appropriate theoretically and best fitting to the Current Population Survey data. Table 4 presents the main results of the reweighting method on the 2012 income reference year. These are used to comment on the size of the nonresponse bias using a stylized difference-in-difference test. Tables 5–6 present our central estimates of the Gini indexes in years 1979–2014 under the reweighting and Pareto replacing correction methods.

Reweighting

Table 4 presents the benchmark results of this study based on household income and the 2012 survey. The main finding is that households' survey response probability is related negatively to income confirming results of Table 1. The coefficients on income $E(\theta_2)$ are consistently negative and highly significant, an indication that unit nonresponse is related to incomes and is therefore expected to bias our measurement of inequality. As a consequence, the corrected Ginis are consistently higher than the non-corrected Ginis. The unweighted corrected Gini coefficient for the 2012 CPS is 49.55 which exceeds the uncorrected Gini by 3.28 percentage points, statistically highly significant. Making use of sampling weights does not affect this finding. The correction for unit nonresponse in this case amounts to 3.35 percentage points of the Gini.

To the extent that applying the statistical agency weights along with the nonresponse-correction weights amounts to some double-correcting for nonresponse and these corrections interact with each other, we can partially correct for this double correction with a Difference-in-Difference (DiD) estimation. The stand-alone correction for nonresponse is estimated at +3.28 percentage points of the Gini (49.55-46.27). The stand-alone correction for non-representative sampling is estimated at +0.59 percentage points of the Gini (46.86-46.27). Adding these effects to the uncorrected Gini (3.28+0.59+46.27), we find that the robust Gini is 50.14 in the CPS. This figure is slightly lower than the original estimate of 50.21, suggesting that the double-correction of nonresponse is responsible for a 0.07 percentage-point inflation of the Gini.

[Table 4]

Applying the same method to the 1979–2014 CPS time series[28] shows that the bias generally increases with income, the uncorrected Gini and the nonresponse rate. Results in Table 5 show that – if we ignore the one insignificant negative correction in year 1995[29] – the estimated bias varies from 0.46% to 6.63% for the non-weighted sample and from 0.42 to 6.83 for the weighted sample, a considerable variation. Changes in the correction bias depend positively on mean income (Pearson=0.53 for the unweighted sample and 0.54 for the weighted sample), nonresponse rate (0.49 and 0.50) and the original uncorrected Gini (0.61 and 0.63). These three variables are also highly correlated among each other. Results can be thought of as "scale effects", meaning that larger values lead to larger biases. However, this is not a law and we can observe years like 2013 and 2014 when income, the Gini and the nonresponse rate are all very high but the nonresponse bias is relatively small. This could be an indication that the quality of the surveys has improved in recent years or a confirmation of the fact that reweighting performs better with lower non-response rates, or both.

[Table 5]

Replacing

Next, we use a methodology first proposed by Cowell and Victoria-Feser (2007) to test sensitivity of the Gini coefficients to extreme or top-coded observations or unit nonresponse on the right-hand side of the income distribution. If top incomes turn out to be influential, we correct for their presence using an estimated Pareto distribution as discussed in the methodological section. If we believe that true top incomes are distributed as Pareto, and that the Pareto coefficient estimates in tables 3 and 6 are near the true value, this would suggest that various top income biases operate in opposite directions. While unit nonresponse and possibly underreporting and top coding of incomes bias the Gini downward, the presence of extreme observations may bias it upward. As a result, the composite correction of the Gini is upward but substantially less that a correction for nonresponse alone.

Results in table 6 show that the bias correction is consistently positive confirming that the CPS estimated Gini is consistently underestimated (estimates for model specifications using other left and right truncation points are available on request). However, the estimated biases are very small ranging from 0.03 to 0.39 for

[28] The CPS is ideal for this exercise as it has been collected systematically for over fifty years, in a consistent format since the 1980s. Income distribution in the CPS has also been consistent across years, with a moderate steady drift in mean incomes and the Gini coefficient.

[29] Year 1995 is unique because the public-access datafile allows us to correctly replicate Census restricted-access statistics on the count of households in each income quantile range for incomes up to $90,000. For all following years, such exact replication is possible only for incomes up to $15,000-35,000, at which point topcoding starts distorting our results. Figure 1 confirms that our public-access data Ginis come closest to the restricted-access data Ginis in year 1995. We conclude that Census topcoding practices and other data properties in that year are responsible for the unexpectedly negative, albeit insignificant, correction.

the non-weighted Gini and from 0.04 to 0.35 for the weighted Gini. Incidentally, the highest correction occurs in years 1993 and 1994 when Census undertook a change in survey design and topcoding. If one believes that the Pareto distribution characterizes observations at the top of the US-CPS distribution well and that extreme observations should not be corrected for, then the Gini suffers from a minor degree of underestimation. In perspective of the findings in preceding sections, the systematic under-representation of top income households due to unit nonresponse is more worrying than other top income issues at biasing inequality estimates systematically downward.

[Table 6]

6. Conclusions

This study has evaluated several methods for correcting statistical problems with top incomes, including unit nonresponse and representativeness of top income observations. The comparative use of two distinct statistical methods for correcting top incomes biases and the sensitivity analysis of their technical specifications were methodological contributions of this study. We first tested for the problem of unit nonresponse by top income households, and corrected for the problem by imputing households' response probability and reweighting them accordingly. We then tested how influential are individual observations at the upper tail of the income distribution, and corrected for the potential problem by replacing actual incomes with values drawn from parametric distributions.

The evidence from the reweighting exercise suggests that unit nonresponse is responsible for a significant 3.3–3.4 percentage point bias in the Gini index of inequality in the 2012 CPS. Across different waves of the CPS, the bias ranges from 0.4 to 6.8 percentage points. The disparities stem from several differences across waves of the CPS including the degree of true inequality, the prevalence of nonresponse, and the geographic incidence of inequality and nonresponse across individual metropolitan statistical areas and states.

Changes in the geographic level of analysis has an important systematic impact on the unit nonresponse correction. Greater degrees of geographic disaggregation typically yield lower estimates of the nonresponse bias, but the nonresponse bias remains significant. The degree of geographic disaggregation is thus an important parameter to consider in correcting for unit nonresponse through reweighting. Understanding of the income distribution, demographics and behavioral similarities in the population within and across regions is important. An experiment on a high quality sample (income reference year 2012) suggested that a medium degree of disaggregation – at the level of states or groups of MCBSAs– achieves the best estimate of the bias and correction for it, compared to extreme aggregation or disaggregation.

Correcting for non-representative distributions of top income observations using fitted values or random draws from the Pareto distribution helps to refine the estimated Gini for the potential presence of various top income issues, which operate in opposite directions. Replacing of top income observations yields a consistent upward correction of the Gini, but this correction is substantially lower than a correction for unit nonresponse alone through reweighting. That may serve as evidence that topmost incomes in the CPS include extreme observations, or that the top 1–10 percent of incomes are distributed more widely than the following 5–25 percent of incomes – on top of our understanding of the extent of nonresponse – would make us believe. This affects inequality measurement in the opposite way as the effect of unit nonresponse. The US Gini corrected in this way rises by 0.03–0.39 percentage points, an insignificant but systematic upward correction across all waves of the CPS.

It should be pointed out that these results, and our conclusions regarding the proper methods of applying the reweighting and replacing methods, were reached on a publicly available version of the US CPS data which suffers from widespread top coding and imputation of incomes. Whether the reweighting, Pareto replacing, and GB2 replacing methods would retain their ordering and their quantitative impacts on the original questionnaire data remains an open question. The Bureau of Labor Statistics should make available the parametric properties of the original data to assist researchers with this determination.

References

Alvaredo, F. and Piketty, T. (2014) Measuring Top Incomes and Inequality in the Middle East: Data Limitations and Illustration with the Case of Egypt. Economic Research Forum, Working Paper 832, May 2014.

An, D. and Little, R.J.A. (2007) Multiple imputation: an alternative to top coding for statistical disclosure control, Journal of the Royal Statistical Society A, 170, 923-940.

Atkinson, A.B. and Micklewright, J. (1983) On the reliability of income data in the family expenditure survey 1970–1977, Journal of the Royal Statistical Society Series A 146, 1, 33-61.

Atkinson, A, Piketty, T. and Saez, E. (2011) Top incomes in the long run of history, Journal of Economic Literature, 49, 3-71.

Bee, C.A., Gathright, G.M.R., Meyer, B.D. (2015) Bias from Unit Non-Response in the Measurement of Income in Household Surveys, August 2015, https://harris.uchicago.edu/files/jsm2015_bgm_unit_non-response_in_cps.pdf.

Bhalla, S. (2002) Imagine there's no country: Poverty, inequality and growth in the era of globalization, Institute for International Economics, Washington, DC.

Bollinger, C.R., Hirsch, B.T., Hokayem, C.M., and Ziliak J.P. (2017) Trouble in the tails? What we know about earnings nonresponse thirty years after Lillard, Smith, and Welch, June 2017, http://www2.gsu.edu/~ecobth/BHHZ_Trouble-in-the-Tails_6-8-2017.pdf.

Bourguignon, F. and Morrisson, C. (2002) Inequality among world citizens: 1820–1992, American Economic Review 92, 4, 727-744.

Burkhauser, R.V., Feng, S. and Larrimore, J. (2010) Improving imputations of top incomes in the public-use current population survey by using both cell-means and variances, Economic Letters, 108, 69-72.

Burkhauser, R.V., Feng, S., Jenkins, S.P. and Larrimore, J. (2011) Estimating trends in US income inequality using the Current Population Survey: the importance of controlling for censoring, Journal of Economic Inequality, 9(3), 393-415.

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.

Census Bureau and Bureau of Labor Statistics, (Census & BLS, 2002) Current Population Survey: Design and methodology, Technical paper 63RV (www.census.gov/prod/2002pubs/tp63rv.pdf), U.S. Government Printing Office, Washington, DC.

Cowell, F.A. and Victoria-Feser, M.-P. (1996) Poverty measurement with contaminated data: A robust approach, European Economic Review, 40, 1761-1771.

Cowell, F.A. and Victoria-Feser, M.-P. (1996) Robustness properties of inequality measures, Econometrica, 64, 77-101.

Cowell, F.A. and Victoria-Feser, M.-P. (2007) Robust Lorenz curves: a semiparametric approach, Journal of Economic Inequality, 5, 21-35.

Cowell, F.A. and Flachaire, E. (2007) Income distribution and inequality measurement: The problem of extreme values, Journal of Econometrics, 141(2), 1044-1072.

Dagum, C. (1980) The generation and distribution of income, the Lorenz curve and the Gini ratio, Economie Appliquee, 33, 327-367.

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), The Review of Economics and Statistics, 87(1), 1-19.

Department of Labor, U.S., and Bureau of Labor Statistics (2015), Consumer Price Index, All Urban Consumers - (CPI-U), US city average, Washington, DC.

Dixon, J. (2007) Nonresponse bias patterns in the Current Population Survey, Bureau of Labor Statistics.

Groves, R.M., and Couper, M. (1998) Nonresponse in Household Interview Surveys, John Wiley and Sons, Inc.: New York, NY.

Hlasny, V. (2016) Unit nonresponse bias to inequality measurement: Worldwide analysis using Luxembourg Income Study database, Luxembourg Income Study technical paper.

Hlasny, V., and Intini, V. (2015) Representativeness of top expenditures in Arab region household surveys, UN-ESCWA/EDID working paper #11, December 2015.

Hlasny, V., and Verme, P. (2016) Top Incomes and the Measurement of Inequality in Egypt, World Bank Economic Review, forthcoming.

Hokayem, C., Raghunathan, T., Rothbaum, J. (2016) Sequential Regression Multivariate Imputation in the Current Population Survey Annual Social and Economic Supplement, September 2016, https://appam.confex.com/appam/2017/webprogram/Paper23281.html.

Jantti, Markus, Tormalehto, Veli-Matti and Marlier, Eric (2013) The Use of Registers in the Context of SILC: Challenges and Opportunities. European Union.

Jenkins, S.P. (2016) Pareto models, top incomes, and recent trends in UK income inequality, ISER Working Paper Series 2016-07.

Jenkins, S.P., Burkhauser, R.V., Feng, S., and Larrimore, J. (2011) Measuring inequality using censored data: a multiple-imputation approach to estimation and inference, Journal of the Royal Statistical Society, 174(1), 63-81.

Jenkins, S.P., and Van Kerm, P. (2007) Paretofit: Stata module fitting a Pareto (Type I) distribution by ML to unit record data, April 2007.

Juster, F.T., and Kuester, K.A. (1991) Differences in the measurement of wealth, wealth inequality and wealth composition obtained from alternative US wealth surveys, Review of Income and Wealth, 37(1), 33-62.

Korinek, A., Mistiaen, J.A. and Ravallion, M. (2006) Survey nonresponse and the distribution of income, Journal of Economic Inequality, 4, 33-55.

Korinek, A., Mistiaen, J.A. and Ravallion, M. (2007) An econometric method of correcting for unit nonresponse bias in surveys, Journal of Econometrics, 136, 213-235.

Lakner, C. and Milanovic, B. (2013) Global income distribution from the fall of the Berlin Wall to the great recession, World Bank Policy Research working paper series #6719.

Larrimore, J., Burkhauser, R.V., Feng, S., Zayatz, L. (2008) Consistent cell means for topcoded incomes in the public use march CPS (1976–2007), Journal of Economic and Social Measurement, 33(2,3), 89-128.

McDonald, J.B. (1984) Some generalized functions for the size distribution of income, Econometrica, 52, 647-663.

Meyer, B.D., Mok, W.K.C. and Sullivan, J.X. (2009) The under-reporting of transfers in household surveys: Its nature and consequences, NBER Working Papers 15181, National Bureau of Economic Research, Inc.

Meyer, B.D. and Mittag, N. (2014) Missclassification in binary choice models. AEA 2014 Proceedings.

Mistiaen, J.A. and Ravallion, M. (2003) Survey compliance and the distribution of income, Policy Research Working Paper #2956, The World Bank.

Modarres, R. and Gastwirth, J. L., (2006), A Cautionary Note on Estimating the Standard Error of the Gini Index of Inequality, Oxford Bulletin of Economics and Statistics, 68, issue 3, p. 385-390.

Pareto, V. (1896) La courbe de la repartition de la richesse, Ecrits sur la courbe de la repartition de la richesse, (writings by Pareto collected by G. Busino, Librairie Droz, 1965), 1-15.

Piketty, T. (2014) Capital in the Twenty-First Century, Belknap Press, USA.

Reiter, J.P. (2003) Inference for partially synthetic, public use microdata sets. Survey Methodology, 29, 181-188.

Sala-i-Martin, X. (2002) The world distribution of income (estimated from individual country distributions), NBER Working Paper No. W8933.

Singh, S.K. and Maddala, G.S. (1976) A function for the size distribution of income, Econometrica, 44, 963-970.

Stiglitz, J. (2011) The Price of Inequality: How Today's Divided Society Endangers Our Future, WW Norton & Company, USA.

Thurow, L.C. (1970) Analyzing the American income distribution, American Economic Review, 60(2), 261-269.

Tiehen, L., Jolliffe, D. and Smeeding, T. (2013) The effect of SNAP on poverty, Brookings Institute Conference paper.

US Census Bureau (2015) Current Population Survey, 2015 Annual Social and Economic (ASEC) Supplement. Wired at: https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar15.pdf.

Table 1. Correction by reweighting for unit nonresponse bias, varying geographic disaggregation

	2012 survey: 24 states	2012 survey: 24 states, households with known MCBSA								
	(1) Analysis at the state level	(2) Census divisions	(3) States	(4) Groups of 10 MCBSAs	(5) Groups of 8 MCBSAs	(6) Groups of 6 MCBSAs	(7) Groups of 4 MCBSAs	(8) Groups of 3 MCBSAs	(9) Groups of 2 MCBSAs	(10) MCBS A level
$E(\theta_1)$	15.024	1.608	14.238	13.894	13.238	12.530	12.350	11.578	11.526	10.316
	(4.532)	(44.008)	(4.932)	(3.678)	(3.218)	(3.768)	(2.998)	(3.214)	(2.905)	(2.652)
$E(\theta_2)$	-1.134	0.010	-1.067	-1.038	-0.982	-0.921	-0.905	-0.838	-0.834	-0.728
	(0.384)	(4.065)	(0.419)	(0.313)	(0.275)	(0.323)	(0.258)	(0.277)	(0.251)	(0.231)
Objective	32.66	605.41	36.86	39.96	41.43	49.34	47.93	57.30	63.44	79.83
Sigma	2.01	63.72	2.13	1.60	1.51	1.74	1.42	1.54	1.44	1.40
Akaike	11.40	35.22	14.30	3.96	0.28	4.34	-17.11	-22.69	-48.91	-151.48
Schwarz	8.78	32.61	11.68	1.34	-2.33	1.73	-19.72	-25.31	-51.52	-154.10
Regions *j*	24 states	7 Census divisions	24 states	40 groups of MCBSAs	45 regions	49 regions	66 regions	80 regions	105 regions	185 MCBSAs
Households *i*	45,596					40,728				
Hhds per region	1,899.8	5,818.3	1,697.0	1,018.2	905.1	831.2	617.1	509.1	387.9	220.2
Uncorrected Gini	47.07					47.27				
	(0.21)					(0.22)				
Gini using sampling wts.	47.18					47.34				
	(0.24)					(0.25)				
Gini corrected for unit nonresp. bias	50.40	47.16	50.16	49.95	49.58	49.20	49.11	48.76	48.73	48.26
	(0.46)	(0.22)	(0.44)	(0.43)	(0.40)	(0.37)	(0.37)	(0.34)	(0.34)	(0.31)
Gini corrected for nonresp. with sampling wts.	**50.63**	**47.26**	**50.35**	**50.14**	**49.76**	**49.38**	**49.29**	**48.92**	**48.90**	**48.41**
	(0.53)	(0.25)	(0.51)	(0.49)	(0.46)	(0.43)	(0.42)	(0.39)	(0.39)	(0.35)
Unit nonresp. bias	3.33	-0.11	2.91	2.68	2.31	1.93	1.84	1.49	1.46	0.99
Bias (using sampling wts.)	**3.45**	**-0.08**	**3.01**	**2.80**	**2.42**	**2.04**	**1.95**	**1.58**	**1.56**	**1.07**

Notes: For clarity, Ginis and their bootstrap standard errors (in parentheses) are multiplied by 100. Ginis in columns 2-3 are also corrected for the state-level inverse rate of MCBSA availability. The 24 states with availability of MCBSA information over 75% of responding and non-responding households include: AZ, CA, CO, CT, DC, DE, FL, GA, LA, MA, MD, IL, MI, NJ, NV, NY, OH, OR, PA, RI, TX, UT, VA, WA.

Table 2. Correction by reweighting for unit nonresponse bias in a trimmed sample

			2012 CPS: 24 states, households with known MCBSA (N=40,728)				
				True uncorrected Gini:	47.34 (.25) ***bias +.00***		
				True Gini using stat. wghts.:	47.01 (.31) ***bias +.00***		
Disaggregation into regions *j*	(1) Uncorrected	(2) 7 Census div.	(3) 24 states	(4) 41 groups of 10 MCBSAs	(5) 61 groups of 4 MCBSAs	(6) 103 groups of 2 MCBSAs	(7) 185 MCBSAs
			5% or 2,036 trimmed, N=38,692				
Gini corrected for unit nonresponse	46.26 ***+1.08****	46.30 ***+1.05****	46.91 ***+.43***	48.65 ***-1.31***	48.52 ***-1.18***	48.14 ***-.80***	46.74 ***+.60****
	(.22)	(.24)	(.28)	(.47)	(.46)	(.42)	(.27)
Gini corrected for unit nonresp. & sampling wghts	45.83 ***+1.18****	45.87 ***+1.14****	46.53 ***+.48***	48.41 ***-1.40***	48.27 ***-1.26***	47.85 ***-.84***	46.34 ***+.67****
	(.27)	(.30)	(.36)	(.62)	(.60)	(.55)	(.34)
			7% or 2,861 trimmed, N=37,867				
Gini corrected for unit nonresponse	46.27 ***+1.07****	46.30 ***+1.04****	46.59 ***+.75****	48.12 ***-.78***	48.05 ***-.71***	47.80 ***-.46***	46.56 ***+.78****
	(.22)	(.24)	(.26)	(.41)	(.41)	(.38)	(.25)
Gini corrected for unit nonresp. & sampling wghts	45.83 ***+1.18****	45.86 ***+1.15****	46.16 ***+.85****	47.79 ***-.78***	47.72 ***-.71***	47.44 ***-.43***	46.13 ***+.88****
	(.27)	(.30)	(.32)	(.53)	(.52)	(.49)	(.31)
			10% or 4,073 trimmed, N=36,655				
Gini corrected for unit nonresponse	46.29 ***+1.05****	46.30 ***+1.04****	46.52 ***+.82****	48.00 ***-.66***	47.87 ***-.53***	47.61 ***-.26***	46.42 ***+.92****
	(.23)	(.25)	(.26)	(.42)	(.40)	(.38)	(.25)
Gini corrected for unit nonresp. & sampling wghts	45.87 ***+1.14****	45.89 ***+1.12****	46.12 ***+.89****	47.72 ***-.71***	47.58 ***-.57***	47.29 ***-.28***	46.01 ***+1.00****
	(.28)	(.31)	(.32)	(.55)	(.53)	(.49)	(.31)
			13% or 5,296 trimmed, N=35,432				
Gini corrected for unit nonresponse	46.32 ***+1.03****	46.34 ***+1.00****	46.51 ***+.83****	47.71 ***-.36***	47.62 ***-.28***	47.38 ***-.03***	46.41 ***+.93****
	(.23)	(.26)	(.26)	(.39)	(.38)	(.36)	(.25)
Gini corrected for unit nonresp. & sampling wghts	45.89 ***+1.12****	45.92 ***+1.09****	46.09 ***+.92****	47.36 ***-.35***	47.26 ***-.25***	47.01 ***+.00***	45.99 ***+1.02****
	(.28)	(.32)	(.32)	(.50)	(.49)	(.46)	(.31)
			16% or 6,517 trimmed, N=34,211				
Gini corrected for unit nonresponse	46.26 ***+1.08****	46.29 ***+1.05****	46.34 ***+1.00****	47.31 ***+.03***	47.24 ***+.10***	47.06 ***+.28***	46.29 ***+1.05****
	(.23)	(.26)	(.25)	(.37)	(.36)	(.34)	(.25)
Gini corrected for unit nonresp. & sampling wghts	45.81 ***+1.20****	45.85 ***+1.16****	45.90 ***+1.11****	46.91 ***+.10***	46.84 ***+.17***	46.65 ***+.36***	45.85 ***+1.16****
	(.29)	(.32)	(.31)	(.47)	(.46)	(.43)	(.31)

Notes: Trimming of observations is randomized, without replacement, subject to household weights given by probability of response (equation 1) where $g=15.0-1.1\,log(income)$. For clarity, Ginis and their standard errors are multiplied by 100. Ginis in all columns are also corrected for the state-level inverse rate of MCBSA availability, to make results comparable to state-wide statistics. Ginis from 30 random draws are computed as per equation 13. Standard errors on Ginis, in parentheses, are bootstrap estimates, and computed as per equation 14. '*' – The remaining bias is significant at 5% (using the standard errors of the true and estimated Ginis, ignoring potential correlation between the two standard errors). The 24 US states with sufficiently high availability of MCBSA information include: AZ, CA, CO, CT, DC, DE, FL, GA, LA, MA, MD, IL, MI, NJ, NV, NY, OH, OR, PA, RI, TX, UT, VA, WA. The 7 US Census divisions are: E.N. Central, Middle Atlantic, Mountain, New England, Pacific, S. Atlantic, W.S. Central.

Table 3. Correction by replacing top incomes with random draws from Pareto distribution

Correction of extreme observations	Sampling correction	Sample size η k obs. replaced	α	β	Gini	Bias in original Gini (pc.pt.)
		Estimation on top 15–h[th] percentile of incomes				
Semi-param. estimation, h=1%	Unweighted	η = 10,321	2.403	1.713	46.27 (.25)	0.00
		k =738	(.036)			
	CPS sampling weights	η = 10,933	2.338	1.747	**46.86 (.21)**	**0.00**
		k = 753	(.040)			
Semi-param. estimation, h=5%	Unweighted	η = 7,372	1.979	2.022	46.36 (.25)	0.09
		k = 3,687	(.090)			
	CPS sampling weights	η = 7,840	1.782	2.278	**47.00 (.34)**	**0.14**
		k = 3,846	(.100)			
Semi-param. estimation, h=8%[i]	Unweighted	η = 5,161	1.280	4.570	47.17 (.46)	0.90
		k = 5,898	(.183)			
	CPS sampling weights	η = 5,450	1.170	6.888	**48.02 (.76)**	**1.16**
		k = 6,236	(.201)			
		Estimation on top 20–h[th] percentile of incomes				
Semi-param. estimation, h=1%	Unweighted	η = 14,007	2.170	1.855	46.28 (.19)	0.01
		k =738	(.027)			
	CPS sampling weights	η = 14,927	2.109	1.901	**46.86 (.27)**	**0.00**
		k = 753	(.030)			
Semi-param. estimation, h=5%	Unweighted	η = 11,058	1.688	2.453	46.45 (.53)	0.18
		k = 3,687	(.056)			
	CPS sampling weights	η = 11,834	1.585	2.710	**47.08 (.96)**	**0.22**
		k = 3,846	(.062)			
Semi-param. estimation, h=10%	Unweighted	η = 7,305	1.503	2.989	47.11 (.56)	0.84
		k = 7,440	(.125)			
	CPS sampling weights	η = 7,936	1.301	4.323	**48.11 (1.32)**	**1.25**
		k = 7,744	(.138)			
		Estimation on top 25–h[th] percentile of incomes				
Semi-param. estimation, h=1%	Unweighted	η = 17,719	2.009	1.991	46.28 (.24)	0.01
		k =738	(.022)			
	CPS sampling weights	η = 18,781	1.908	2.102	**46.87 (.27)**	**0.01**
		k =753	(.024)			
Semi-param. estimation, h=5%	Unweighted	η = 14,770	1.562	2.778	46.51 (.77)	0.24
		k = 3,687	(.040)			
	CPS sampling weights	η = 15,688	1.391	3.557	**47.19 (1.25)**	**0.33**
		k = 3,846	(.044)			
Semi-param. estimation, h=10%	Unweighted	η = 11,017	1.416	3.403	47.27 (1.72)	1.00
		k = 7,440	(.074)			
	CPS sampling weights	η = 11,790	1.103	10.696	**48.79 (1.96)**	**1.93**
		k = 7,744	(.081)			
Sample size (households)			73,723			

Notes: Pareto coefficients are estimated on non-contaminated income observations (sample size $\eta; L \leq x < H; H$ is income corresponding to the 100-h[th] percentile) using maximum likelihood. Semi-parametric Gini coefficients are computed as in equations 7 and 8. Their standard errors, in parentheses, are jackknife estimates and are computed using 30 random draws from the estimated Pareto distribution as in equation 14. Unit nonresponse bias is corrected using geographic disaggregation at the level of EU subnational regions, and US states. For clarity, Ginis and their standard errors are multiplied by 100.

i This right-truncation is at a higher level than in the analyses below, to keep the estimation sample (range of income quantiles on which Pareto distribution is fit) large enough compared to the prediction sample (quantiles for which Pareto estimates are drawn).

Table 4. Benchmark results of Gini correction by reweighting for unit nonresponse bias

Income reference year	2012
$E(\theta_1)$	14.889 (3.029)
$E(\theta_2)$	-1.121 (0.258)
Regions *j*	51 states
Households *i*	73,746
Uncorrected Gini	46.27 (0.17)
Gini using stat. agency weights	46.86 (0.20)
Gini corrected for unit nonresponse bias	49.55 (0.35)
Gini corrected for unit non-resp. bias, with sampling wts.	**50.21** (0.42)
Unit nonresponse bias	3.28
Bias (using sampling wts.)	**3.35**

The model is estimated on an unweighted sample, and the uncorrected or corrected weights are only applied in the calculation of the Ginis. Standard errors are in parentheses. Ginis and their bootstrap standard errors are multiplied by 100. Only incomes ≥1 are retained.

Table 5. Correction by reweighting for unit nonresponse bias, CPS income years 1979–2014

Year	Mean income ($)	Nonresp. rate (%)	Gini, non-weighted	Gini, CPS-wtd. hhds	$E(\theta_1)$	$E(\theta_2)$	Gini, nonresp. corrected	Gini, non-resp. & CPS wghted.	Unit nonresp. bias in Gini	Bias in CPS weighted Gini
1979	19,390.88	4.34	38.97 (.10)	39.05 (.12)	25.62 (3.61)	-2.23 (.34)	39.92 (.16)	**40.05 (.21)**	0.95	**1.00**
1980	20,851.11	4.37	38.88 (.10)	38.89 (.12)	23.54 (3.31)	-2.01 (.31)	39.59 (.21)	**39.64 (.25)**	0.71	**0.75**
1981	22,834.67	4.61	39.69 (.11)	39.79 (.13)	23.13 (5.56)	-1.96 (.52)	40.42 (.13)	**40.56 (.16)**	0.73	**0.77**
1982	24,339.96	4.48	40.28 (.11)	40.35 (.13)	20.89 (4.41)	-1.74 (.41)	40.92 (.14)	**41.00 (.16)**	0.64	**0.65**
1983	25,397.26	4.99	40.39 (.11)	40.49 (.12)	19.54 (6.28)	-1.61 (.59)	40.85 (.13)	**40.96 (.15)**	0.46	**0.47**
1984	27,641.77	5.24	40.96 (.12)	41.10 (.13)	24.66 (4.58)	-2.07 (.42)	42.64 (.20)	**42.87 (.24)**	1.68	**1.77**
1985	29,006.68	5.66	41.02 (.12)	41.04 (.13)	19.63 (5.69)	-1.61 (.52)	41.76 (.15)	**41.72 (.15)**	0.74	**0.68**
1986	30,599.29	5.50	41.34 (.12)	41.39 (.13)	22.69 (3.96)	-1.88 (.36)	42.49 (.17)	**42.56 (.19)**	1.15	**1.17**
1987	32,125.49	5.30	41.39 (.12)	41.49 (.13)	27.48 (2.77)	-2.30 (.25)	43.39 (.29)	**43.46 (.27)**	2.00	**1.97**
1988	33,620.47	5.47	41.34 (.12)	41.45 (.14)	29.44 (3.03)	-2.47 (.27)	43.44 (.25)	**43.35 (.22)**	2.10	**1.90**
1989	35,864.26	4.56	41.50 (.11)	41.55 (.13)	24.74 (4.42)	-2.02 (.39)	42.66 (.15)	**42.71 (.17)**	1.16	**1.16**
1990	36,804.27	4.71	41.36 (.11)	41.38 (.13)	24.02 (3.85)	-1.96 (.34)	42.29 (.14)	**42.33 (.16)**	0.93	**0.95**
1991	37,497.71	5.03	41.50 (.11)	41.46 (.13)	22.48 (2.78)	-1.82 (.25)	42.40 (.15)	**42.29 (.16)**	0.90	**0.83**
1992	38,351.00	5.00	41.76 (.11)	41.81 (.13)	21.76 (3.72)	-1.75 (.33)	42.49 (.14)	**42.50 (.16)**	0.73	**0.69**
1993	39,637.70	7.27	42.17 (.12)	42.11 (.13)	17.26 (3.25)	-1.38 (.29)	42.66 (.14)	**42.53 (.16)**	0.49	**0.42**
1994	41,174.80	7.02	42.26 (.11)	42.25 (.13)	7.85 (3.94)	-0.50 (.37)	42.11 (.11)	**42.09 (.13)**	-0.15	**-0.16**
1995	45,220.38	7.70	44.59 (.17)	44.63 (.20)	14.60 (4.20)	-1.13 (.38)	46.21 (.23)	**46.31 (.26)**	1.62	**1.68**
1996	47,454.13	7.18	44.98 (.17)	45.05 (.19)	16.92 (3.14)	-1.32 (.28)	47.39 (.24)	**47.49 (.27)**	2.41	**2.44**
1997	50,104.64	7.80	45.38 (.17)	45.32 (.19)	22.30 (3.15)	-1.79 (.27)	50.34 (.32)	**50.22 (.34)**	4.96	**4.90**
1998	52,326.19	7.91	45.02 (.17)	44.98 (.19)	20.51 (3.31)	-1.63 (.28)	48.83 (.27)	**48.73 (.30)**	3.81	**3.75**
1999	54,219.48	6.89	44.05 (.14)	44.13 (.15)	24.62 (3.57)	-1.97 (.30)	46.85 (.23)	**46.77 (.23)**	2.80	**2.64**
2000	57,488.72	8.03	45.08 (.17)	45.42 (.20)	23.74 (2.55)	-1.89 (.21)	49.73 (.27)	**50.21 (.34)**	4.65	**4.79**
2001	58,807.79	7.81	44.96 (.13)	45.87 (.16)	21.24 (3.12)	-1.67 (.26)	48.57 (.20)	**49.49 (.23)**	3.61	**3.62**
2002	58,467.77	8.05	44.84 (.14)	45.56 (.16)	21.75 (2.82)	-1.72 (.24)	49.48 (.25)	**50.16 (.29)**	4.64	**4.60**
2003	59,800.95	8.37	44.95 (.14)	45.67 (.16)	18.46 (3.39)	-1.44 (.29)	47.89 (.21)	**48.60 (.26)**	2.94	**2.93**
2004	61,250.15	9.02	45.04 (.14)	45.86 (.17)	24.73 (2.74)	-1.97 (.23)	51.67 (.33)	**52.69 (.44)**	6.63	**6.83**
2005	64,149.54	8.61	45.46 (.14)	46.12 (.16)	25.84 (3.33)	-2.04 (.27)	51.88 (.29)	**52.57 (.35)**	6.42	**6.45**
2006	67,420.38	8.67	45.53 (.14)	46.19 (.16)	19.41 (2.99)	-1.50 (.25)	49.37 (.23)	**50.00 (.26)**	3.84	**3.81**
2007	68,487.02	7.82	44.79 (.13)	45.48 (.15)	19.86 (3.06)	-1.53 (.26)	47.68 (.20)	**48.33 (.23)**	2.89	**2.85**
2008	69,308.99	7.06	45.04 (.13)	45.77 (.15)	20.13 (3.02)	-1.54 (.25)	47.59 (.18)	**48.31 (.21)**	2.55	**2.54**
2009	68,909.96	7.01	45.39 (.13)	45.93 (.16)	18.05 (2.86)	-1.37 (.24)	47.58 (.19)	**48.13 (.21)**	2.19	**2.20**
2010	68,591.28	8.12	45.40 (.15)	46.07 (.18)	13.47 (3.06)	-0.99 (.27)	46.89 (.23)	**47.64 (.28)**	1.49	**1.57**
2011	70,781.66	8.93	46.21 (.16)	46.91 (.20)	15.65 (2.55)	-1.18 (.22)	49.49 (.31)	**50.35 (.37)**	3.28	**3.44**
2012	72,353.13	9.54	46.27 (.17)	46.86 (.20)	14.89 (3.03)	-1.12 (.26)	49.55 (.35)	**50.21 (.42)**	3.28	**3.35**
2013t	73,721.17	10.26	46.33 (.19)	46.86 (.23)	13.11 (4.46)	-0.97 (.38)	48.27 (.29)	**48.95 (.35)**	1.94	**2.09**
2013r	76,553.83	10.57	46.65 (.30)	47.43 (.36)	9.99 (4.14)	-0.70 (.36)	47.60 (.40)	**48.51 (.49)**	0.95	**1.08**
2014	76,921.34	12.33	46.81 (.16)	47.20 (.17)	11.76 (4.19)	-0.87 (.36)	48.48 (.24)	**48.84 (.26)**	1.67	**1.64**

Notes: For clarity, Ginis and their bootstrap standard errors (in parentheses) are multiplied by 100.

Table 6. Correction by replacing top incomes with random draws from Pareto distribution, CPS income years 1979–2014

Year	Pareto α, CPS non-wghted. sample	Pareto β, CPS non-wghted.	Semiparam. Gini, non-wghted.	Pareto α, CPS wghted. sample	Pareto β, CPS wghted.	Semiparam. Gini, CPS wghted.	Bias in non-wghted. Gini	Bias in CPS wghted. Gini
1979	2.21 (0.12)	1.82	39.11 (0.33)	2.15 (0.14)	1.87	**39.20 (0.26)**	0.14	**0.15**
1980	2.18 (0.12)	1.85	39.03 (0.21)	1.96 (0.14)	2.04	**39.07 (0.73)**	0.14	**0.18**
1981	2.12 (0.13)	1.89	39.84 (0.24)	1.94 (0.14)	2.06	**39.97 (0.31)**	0.15	**0.18**
1982	2.46 (0.12)	1.69	40.39 (0.18)	2.29 (0.14)	1.78	**40.47 (0.48)**	0.11	**0.13**
1983	2.25 (0.12)	1.80	40.53 (0.25)	2.05 (0.13)	1.95	**40.65 (0.35)**	0.14	**0.17**
1984	2.34 (0.12)	1.74	41.07 (0.25)	2.40 (0.13)	1.71	**41.21 (0.24)**	0.11	**0.10**
1985	2.56 (0.12)	1.64	41.11 (0.26)	2.65 (0.14)	1.60	**41.12 (0.19)**	0.09	**0.09**
1986	2.32 (0.12)	1.76	41.46 (0.21)	2.34 (0.13)	1.75	**41.51 (0.30)**	0.12	**0.12**
1987	2.36 (0.12)	1.74	41.50 (0.21)	2.48 (0.13)	1.68	**41.59 (0.26)**	0.12	**0.10**
1988	2.31 (0.12)	1.76	41.46 (0.29)	2.33 (0.14)	1.75	**41.57 (0.25)**	0.13	**0.13**
1989	2.23 (0.11)	1.81	41.63 (0.22)	2.11 (0.13)	1.90	**41.70 (0.30)**	0.14	**0.15**
1990	2.12 (0.11)	1.89	41.52 (0.57)	2.11 (0.13)	1.90	**41.54 (0.29)**	0.16	**0.16**
1991	2.17 (0.11)	1.86	41.65 (0.31)	2.18 (0.13)	1.85	**41.61 (0.30)**	0.15	**0.15**
1992	2.05 (0.11)	1.95	41.93 (0.25)	1.95 (0.13)	2.06	**42.00 (0.74)**	0.17	**0.19**
1993	1.62 (0.12)	2.61	42.44 (0.49)	1.51 (0.14)	2.97	**42.42 (0.91)**	0.27	**0.31**
1994	1.36 (0.12)	3.75	42.65 (0.93)	1.44 (0.14)	3.26	**42.59 (0.99)**	0.39	**0.35**
1995	2.11 (0.12)	1.90	44.67 (0.29)	2.16 (0.14)	1.86	**44.70 (0.37)**	0.08	**0.07**
1996	2.01 (0.11)	1.99	45.07 (0.28)	1.98 (0.13)	2.02	**45.15 (0.33)**	0.09	**0.10**
1997	2.06 (0.11)	1.94	45.46 (0.31)	1.98 (0.13)	2.02	**45.43 (0.36)**	0.09	**0.11**
1998	2.05 (0.11)	1.95	45.11 (0.30)	1.99 (0.13)	2.01	**45.09 (0.31)**	0.09	**0.11**
1999	2.27 (0.11)	1.79	44.17 (0.77)	2.14 (0.13)	1.88	**44.27 (0.26)**	0.11	**0.14**
2000	2.25 (0.11)	1.80	45.15 (0.23)	2.24 (0.13)	1.81	**45.49 (0.31)**	0.07	**0.08**
2001	2.40 (0.09)	1.72	45.01 (0.19)	2.25 (0.11)	1.80	**45.95 (0.54)**	0.05	**0.08**
2002	2.39 (0.09)	1.72	44.87 (0.25)	2.34 (0.11)	1.74	**45.59 (0.24)**	0.03	**0.04**
2003	2.00 (0.09)	2.00	45.06 (0.28)	1.97 (0.11)	2.03	**45.79 (0.39)**	0.11	**0.12**
2004	1.99 (0.09)	2.01	45.14 (0.45)	1.68 (0.11)	2.48	**46.06 (0.54)**	0.10	**0.20**
2005	1.96 (0.09)	2.04	45.58 (0.29)	1.84 (0.10)	2.19	**46.27 (0.42)**	0.12	**0.15**
2006	1.97 (0.09)	2.04	45.64 (0.57)	1.76 (0.10)	2.32	**46.36 (0.42)**	0.11	**0.17**
2007	2.17 (0.09)	1.85	44.87 (0.21)	1.90 (0.11)	2.12	**45.62 (0.38)**	0.08	**0.15**
2008	1.97 (0.10)	2.03	45.17 (0.28)	1.74 (0.11)	2.35	**45.97 (0.32)**	0.13	**0.20**
2009	1.94 (0.09)	2.07	45.52 (0.24)	1.83 (0.11)	2.20	**46.09 (0.61)**	0.13	**0.17**
2010	1.94 (0.09)	2.06	45.53 (0.42)	1.93 (0.11)	2.08	**46.20 (0.42)**	0.13	**0.13**
2011	1.91 (0.09)	2.10	46.32 (0.38)	1.82 (0.10)	2.22	**47.05 (0.71)**	0.11	**0.14**
2012	1.98 (0.09)	2.02	46.36 (0.25)	1.78 (0.10)	2.28	**47.00 (0.62)**	0.09	**0.15**
2013t	2.01 (0.11)	1.99	46.42 (0.30)	1.90 (0.12)	2.11	**46.98 (0.49)**	0.10	**0.12**
2013r	1.95 (0.17)	2.05	46.76 (0.61)	1.53 (0.18)	2.87	**47.68 (0.90)**	0.10	**0.25**
2014	1.84 (0.09)	2.19	46.95 (0.53)	1.61 (0.10)	2.63	**47.43 (0.69)**	0.14	**0.23**

Notes: The lower bound for estimation of Pareto distribution is the 85th income percentile. The cutoff for replacement with random draws from Pareto distribution is the 95th income percentile. For clarity, Ginis and their bootstrap standard errors (in parentheses) are multiplied by 100.

Annex

Table A1. Nonresponse rate and income distribution by state, CPS income year 2012

State	Metrop. CBSAs	MCBSA Known (% hhds.)	Responding Households	State Non-resp. Rate (%)	Mean Income ($)	Gini, CPS-wtd. hhds
Alabama	8	71.2	818	6.2	62,700.29	47.08 (2.06)
Alaska	0	0.0	859	12.7	83,332.89	42.80 (1.79)
Arizona	3	84.7	934	8.4	66,119.68	47.06 (1.76)
Arkansas	3	49.5	826	5.6	56,110.07	47.25 (1.69)
California	23	98.4	6,747	8.6	83,372.45	48.29 (0.59)
Colorado	6	89.0	1,646	9.2	80,259.79	45.23 (0.89)
Connecticut	6	92.1	1,592	12.5	91,914.71	48.22 (1.10)
Delaware	2	79.8	1,134	8.2	67,089.25	43.20 (1.15)
Distr. Columbia	1	100.0	1,297	13.3	98,508.22	51.40 (1.55)
Florida	19	96.2	3,136	5.1	65,073.03	46.03 (0.89)
Georgia	10	82.0	1,608	6.9	67,494.40	45.57 (1.12)
Hawaii	1	70.2	1,215	6.8	76,991.34	45.83 (1.42)
Idaho	2	46.7	767	8.9	63,068.17	43.63 (1.48)
Illinois	10	89.1	2,240	8.3	78,029.28	49.06 (1.23)
Indiana	9	69.7	1,091	8.3	63,553.75	43.49 (0.96)
Iowa	6	49.1	1,361	7.1	67,676.32	41.65 (1.05)
Kansas	4	65.1	1,049	8.8	68,919.63	44.96 (1.66)
Kentucky	4	48.0	1,031	8.2	55,118.69	42.32 (0.90)
Louisiana	6	82.9	754	7.4	59,170.90	45.82 (1.06)
Maine	2	40.3	1,172	13.4	65,221.48	43.08 (0.90)
Maryland	4	92.7	1,736	15.4	91,454.12	43.54 (1.00)
Massachusetts	6	94.1	1,070	12.9	85,106.21	46.53 (1.20)
Michigan	12	83.4	1,636	9.8	69,661.07	46.07 (1.30)
Minnesota	3	70.2	1,706	9.1	80,135.57	43.33 (1.08)
Mississippi	3	33.4	712	8.1	55,434.66	48.49 (2.20)
Missouri	5	70.5	1,151	8.4	69,696.19	47.69 (1.63)
Montana	1	12.5	707	4.1	58,137.05	42.36 (1.52)
Nebraska	1	40.8	1,104	9.3	68,507.85	41.20 (0.95)
Nevada	2	87.7	1,147	10.4	65,124.21	43.97 (1.12)
New Hampshire	2	41.9	1,402	12.5	85,360.26	42.40 (1.15)
New Jersey	7	100.0	1,412	13.5	90,940.01	46.94 (1.03)
New Mexico	4	69.7	726	7.6	76,913.35	54.75 (2.58)
New York	9	92.3	3,143	13.9	74,257.71	50.34 (0.97)
North Carolina	9	64.3	1,520	8.7	59,262.14	46.20 (1.00)
North Dakota	1	26.7	922	6.9	76,530.98	45.70 (1.63)
Ohio	9	75.7	1,961	10.2	61,996.53	45.29 (0.90)
Oklahoma	3	67.6	906	7.0	64,982.57	44.76 (1.72)
Oregon	5	76.4	1,012	11.8	66,318.12	42.24 (1.05)
Pennsylvania	11	82.6	2,197	9.6	69,869.10	45.16 (0.89)
Rhode Island	1	100.0	1,192	15.3	81,879.41	49.10 (1.23)
South Carolina	8	66.8	1,016	6.2	58,455.64	43.79 (1.00)
South Dakota	1	27.2	1,065	8.0	65,696.04	44.32 (1.69)
Tennessee	6	65.8	1,003	8.7	59,604.70	46.01 (1.40)
Texas	17	86.4	4,310	9.8	72,255.19	46.12 (0.84)
Utah	3	77.9	861	6.9	74,783.71	42.54 (1.32)
Vermont	1	32.4	964	14.8	71,790.66	42.91 (1.23)
Virginia	6	82.4	1,568	8.8	85,722.53	45.69 (1.15)
Washington	7	83.1	1,283	9.8	82,437.54	44.48 (1.25)
West Virginia	2	28.6	716	6.6	60,239.76	46.77 (2.28)
Wisconsin	10	69.6	1,405	6.7	70,119.60	43.87 (1.16)
Wyoming	0	0.0	935	9.7	71,335.46	41.83 (1.04)
Wtd. Mean [Total]	5.57 [284]	74.6	1,446 [73,765]	9.5	72,353.13	46.86 (0.20)

Notes: MCBSA availability is reported for both responding and non-responding households. Nonresponse rate is reported in the survey at the state level (and is available also at the level of MCBSAs and counties for 74.6% and 43.0% of households, respectively). Per-capita income is weighted by household size. Mean incomes may not be representative of those for the entire states, as they omit non-responding households. For clarity, Ginis are multiplied by 100.

Table A2. Correction by replacing top incomes with random draws from GB2 distribution

Correction of extreme obs.	Sample size η k obs. replaced	log(a)	log(b)	log(p)	log(q)	Gini	Bias in original Gini
Semi-param. estimation, h=1%	η = 52,073 k = 752	0.29 (.05)	12.43 (.03)	0.12 (.14)	1.86 (.05)	**44.70 (.12)**	**-2.16**
Semi-param. estimation, h=5%	η = 48,980 k = 3,845	1.69 (.00)	14.93 (.00)	-3.86 (.11)	15.76 (.00)	**39.80 (.09)**	**-7.06**
Semi-param. estimation, h=10%	η = 45,094 k = 7,731	-2.20 (.00)	10.73 (--)	6.54 (--)	6.51 (.00)	**39.67 (.10)**	**-7.19**

Notes: Sample size 73,723 households. CPS sampling weights used in estimation. Huber-White robust standard errors in parentheses. '--' Standard error on some parameters in some runs could not be estimated.