Top Incomes and the Measurement of Inequality: A Comparative Analysis of Correction Methods using EU, US and Egyptian Survey Data

Vladimir Hlasny (Ewha Womans University, Seoul and UN-ESCWA)
Paolo Verme (World Bank)

Paper prepared for the IARIW-CAPMAS Special Conference "Experiences and Challenges in Measuring Income, Wealth, Poverty and Inequality in the Middle East and North Africa"

Cairo, Egypt
November 23-25, 2015

Session 7: Inequality II
Tuesday, November 24, 2015
12:45-14:45

Top Incomes and the Measurement of Inequality: A Comparative Analysis of Correction Methods using EU, US and Egyptian Survey Data

Vladimir Hlasny and Paolo Verme

Abstract

It is sometimes observed and frequently assumed that top incomes in household surveys worldwide are poorly measured and that this problem biases the measurement of income inequality. This paper tests this assumption and compares the performance of *reweighting* and *replacing* methods designed to correct inequality measures for income biases generated by unit or item non-response. The European Union's Statistics on Income and Living Conditions (EU-SILC), the United States' Current Population Survey (US-CPS) and the Egyptian Household Income, Expenditure and Consumption Survey (EG-HIECS) are used as prototypes of vastly different data sets. Results indicate that survey response probability is negatively associated to income per capita thereby confirming that unit or item non-response bias the measurement of income inequality. When using *reweighting* methods, the higher the level of geographical disaggregation the lower the estimated bias of the Gini. Middle levels of geographical disaggregation are found to perform better than hyper aggregation or hyper disaggregation. When using *replacing* methods, the Pareto coefficient is sensitive to the cut-off point of the distribution. The use of Pareto distributions results in larger corrections as compared to the use of generalized beta distributions but the difference is not very large.

JEL: D31, D63, N35.

Keywords: Top incomes, inequality measures, survey non-response, Pareto distribution, parametric estimation.

Sector Board: Poverty (POV)

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. Economics research had somehow anticipated this interest with the emergence of a body of literature that focused on the long-term evolution of top incomes as nicely summarized in Atkinson et al. (2011). Thanks to these studies and the wider public attention that top incomes have received in the aftermath of the global financial crisis, it is now 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, although proper data on top incomes remain limited to a handful of countries

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, and Davidson and Flachaire, 2007) and trends in incomes of the richest 1% of households have been driving 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. Hence, 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 precision of reported 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 non-response, item non-response, item underreporting and other measurement errors), data preparation (top coding trimming or censoring, provision of subsamples) or data analysis (trimming of outliers, choice of estimator). Even the most sophisticated surveys such as the Current Population Survey (CPS) – the official source for income, inequality and poverty estimation 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 non-response particularly by high-income households (Korinek et al. 2006; Dixon 2007).[1] Poor income measurement can also explain differences in inequality measurements 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 non-response, and sample attrition rates.

There are essentially 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 between macro and micro data. We can call this class of methods the *out of surveys correction methods*. Burkhauser et al. (2012) report that tax-record and income-survey data may yield different measures of

[1] The U.S. Census Bureau provides a limited correction for unit non-response by reweighting 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.

income inequality because of differences in income components and different definitions of inequality. Deaton (2005) shows how unit non-response 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). 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. Lakner and Milanovic (2013) proposed another approach by combining corrections for unit non-response with corrections for measurement errors among top incomes and calibrate the estimated Pareto distribution among top incomes using aggregate income information from national accounts data. This method essentially assigns any disparity between the national accounts and household surveys to top-income households, effectively accounting for both unit non-response and measurement errors.

The second class of methods focuses instead on micro 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 non-response rates across geographical areas. This method is used to correct for unit non-response (Mistiaen and Ravallion 2003; Korinek et al. 2006 and 2007) but can also be applied to item non-response. 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 non-responses if these non-responses are concentrated among top incomes (Cowell and Victoria-Feser, 2007; Jenkins et al. 2011)

This paper follows this last class of methods by testing the reweighting and replacing methods in the presence of different types of data. It is evident that both classes of methods have their advantages and disadvantages. Using tax records to inform the measurement of top incomes has its own measurement problems while the information available within surveys has its limits even if used creatively to correct for top incomes. However, the *out of survey correction methods* have a strong limitation in that good tax or macro data are only available in wealthy and largely formal economies and cannot be applied to most of the developing world, while household survey data of reasonable quality are now available in most of the developing world.

The paper is organized as follows. The next section discusses measurement issues related to top incomes. The following section outlines the main methods used to correct for top income biases related to unit non-response. Section four describes the data. Section five presents the main results and section six concludes.

Measurement issues

Problems related to top-income data may be due to sample design, data collection, data preparation or data analysis. We introduce these four typologies of errors in turn.

<u>Sample design issues</u> emerge when the sampling is designed in such a way that top incomes cannot be captured by design. This can occur, for example, when the sampling is done poorly or when the

population census is old or the master sample has not been updated to capture newly constructed wealthy areas. If detected, some of these issues can be corrected post-survey by reweighting the sample, but either detecting or correcting these problems post-survey is not simple. To note here that we should not expect exceptionally high incomes to be captured in household sample surveys. Billionaires are a very rare characteristic in any population. There are less than 3,000 people worldwide with this characteristic and most countries have only one or two billionaires at the most. If one wishes to study billionaires, sample surveys are not the right instrument. It would also be unwise to add billionaires in survey income statistics partly because they are billionaires in wealth, not income, and partly because most of their wealth is generated globally rather than in a particular country. Including billionaires in income statistics would simply bias survey population statistics. Therefore, when we consider the very top income earners in this paper we are considering millionaires in wealth whose income is counted in the hundreds of thousands dollars annually. This is the class of people we want properly represented in household sample surveys at the top of the distribution.

Data collection issues mostly arise from respondents' or interviewers' non-compliance to survey instructions and may result in unit non-response, item non-response, item underreporting or generic measurement errors:

Unit non-response. Unit non-response refers to households that were selected into the sample but did not participate in the survey. The reasons for non-participation can be many such as a change of address or non-interest on the part of the household. Interviewers generally have lists of addresses that can be used to replace the missing household but this practice is not always sufficient to complete the survey with the full expected sample. Most of the available household survey data suffer from unit non-response. In some surveys, the reason for non-response is recorded but in others it is not. Unit non-response bias results if non-response is not random but systematically driven by specific factors. This paper will address unit non-response issues.

Item non-response. Item non-response occurs when households participating in the survey do not reply to an item of interest (income or expenditure in our case). Item non-responses may be related to households' characteristics such as wealth or education, and this may bias statistics constructed with income or expenditure variables. Item non-response biases results if non-response is not random and is related to specific factors. As compared to unit non-response, it is possible to correct for item non-response using information on the reasons for non-response (when available). The methods proposed in this paper for unit non-response also apply to item non-response. The only difference is that with item non-response one can also use alternative correction methods such as imputation techniques which are not available for unit non-response.

Item underreporting. Consistent underreporting of variables on the part of respondents can lead to poor estimates of inequality. For example, if the degree of underreporting rises with income, the measurement of inequality could be affected. Even if underreporting applies equally across respondents, the measurement of inequality may change if the income inequality measure used is not scale invariant. Over-reporting is also possible although extremely rare with income and expenditure data, particularly at the top end of the distribution. Item underreporting will not be treated in this paper explicitly but one of the methodologies proposed will implicitly consider this issue.

Generic measurement errors. Any variable including income or expenditure can be subject to measurement error. This error is typically expected to be random, distributed approximately normally and with zero mean. For example, extreme observations in an income distribution can result from data input errors, but if they are very large they bias sample statistics significantly. Statistical agencies are usually quite thorough on this issue and clear data of errors before providing the data to researchers. This issue will not be treated in this paper.

<u>Data preparation issues</u> are mostly a consequence of statistical agencies' compliance with rules and regulations governing data confidentiality and data use, and may result in top coding, sample trimming, or the provision of limited subsamples to researchers.

Top coding. Top coding is the practice adopted by some statistical agencies to modify intentionally the values of some variables to prevent identification of households or individuals. Many agencies replace values above a certain threshold level with the minimum or mean of the variable in a group (cell) of similar units. In recent waves of the US-CPS, top coding is conducted via "rank proximity swapping," whereas values above the cutoff for top-coding are swapped within demographic cells for another value within bounded intervals, and then rounded off. The imputed values are thus distributed similarly as the latent true values, but individual values are not identical to the true values and generate statistical errors. In some cases and for research purposes, statistical agencies provide restricted access to the original values. But in most cases researchers are left with the problem of having to correct sample statistics for top coding. Correcting for top coding is similar to addressing the issues of data censoring. In this paper, we will not review techniques to work with censored data but, rather, use and compare the performance of methods that replace potentially censored or top coded observations.

Trimming. Trimming is the practice of cutting off some observations from the sample. This may be done for confidentiality reasons or for observations that appear unreliable. Researchers may not be informed whether statistical agencies have trimmed data, why trimming was performed, or both. A related issue is that of trimming through sampling weights. Statistical agencies sometimes trim sampling weights to bring them within a narrow range of values. The objective is to limit the influence of units that are rare in the sampling frame, particularly if their variable values may have been mismeasured. Trimming observations or weights biases statistical measurement and should be corrected for. The replacing methods evaluated in this paper address the problem of trimming observations at the top..

Provision of subsamples. Some statistical agencies cannot provide the entire data sets to researchers for confidentiality or national-security reasons or simply to prevent others from replicating official statistics. In many countries, statistical agencies provide 20% to 50% of their samples to researchers. These subsamples are usually extracted randomly so that statistics produced from these subsamples may be reasonably accurate. As we know from sampling theory, random extraction is the best option for extracting a subsample in the absence of any information on the underlying population. However, only one subsample is typically extracted from the full sample and given to researchers and this implies that a particularly "unlucky" random extraction can potentially provide skewed estimates of the statistics of interest. Hlasny and Verme (2014) have tested the margins of errors in inequality measurement that can arise from the provision of subsamples instead of full samples and found significant margins of errors.

<u>Data analysis issues</u> may arise from an inadvertently wrong choice of statistical estimators on the part of researchers. Some estimators are more sensitive than others to the issues listed above so that one choice of

estimator may lead to greater errors than others. For example, Cowell and Victoria-Feser (1996) have found that the Gini index is more robust to contamination of extreme values than two members of the generalized entropy family, a finding later confirmed by Cowell and Flachaire (2007). In what follows, we will focus on the Gini index and leave the discussion of alternative inequality estimators aside. Also important to note is that many researchers routinely trim outliers or problematic observations or apply top coding with little consideration for the implications for the measurement of inequality.

2. Models

As for the discussion in the previous section, this paper will cover and compare techniques that can be used to correct inequality estimates for top income issues related to unit non-response, item non-response, 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 non-response. Unlike in the case of item non-response, 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 non-response rates across regions, whereas the mass of respondents in a region is 'grossed up' uniformly by the regional non-response 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 non-response 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 non-response rates across geographical units as well as information about the distribution within units.[2] It is assumed that the probability of a household *i* to respond to the survey, P_i, is a logistic function of its arguments:

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

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

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

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

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).

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

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 non-response bias in the observed distribution of the demographic 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 $\hat{m}_j$ 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 non-response 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, potentially calling for larger geographic units.

Properties of the data at hand appear to 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 level of sample stratification are the parameters to consider. Korinek et al. (2006, 2007) used state-level disaggregation of US CPS data, because geographic identifiers are consistently reported only at that level whereas county or metropolitan statistical area identifiers are missing for a large portion of responding as well as non-responding households. In their analysis of the Egyptian Household Income, Expenditure and Consumption Survey (HIECS), Hlasny and Verme (2013) considered regional disaggregation both at the highest administrative level (governorate by urban–rural areas, 50 areas with 939.7 observations on average) and at the level of primary sampling units (PSUs, 2,526 areas with 18.6 average observations). 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 also 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 *j* areas, 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 non-response bias across the alternative specifications, and to identify the preferred degree of disaggregation for various types of data.

Finally worth noting, to satisfy the assumption of stability of $g(x_i,\theta)$, the geographic extent of analysis should be limited to regions in which households are behaviorally similar, in the sense that households with similar values of demographic variables are expected to have a similar response probability across all regions. On the margins we will report how the exclusion of influential regions affects the correction for the unit non-response bias.

Functional form. The relationship between households' characteristics and their response probability can be modeled in a number of ways including logit or probit functions. This paper uses logit for modeling convenience and in deference to previous literature. Furthermore, equation 1 allows various functional forms of household characteristics. Korinek et al. (2006, 2007) and Hlasny and Verme (2013) evaluated specifications with varying degrees of allowed curvature, with or without monotonicity. They concluded that logarithmic specification of income yields better fit than linear, quadratic or higher-order polynomial forms, implying that unit non-response problem is concentrated in one end of the income distribution. This paper takes the problem of functional form as settled, and uses logistic probability model and logarithmic functional form in equation 1.

Explanatory variables. Korinek et al. (2006, 2007) evaluated a number of variables affecting households' response probability, including income, gender, race, age, education, employment status, household size and an urban–rural indicator. Hlasny and Verme (2013) compared income and expenditures. The choice over covariates involves a tradeoff between robustness of complete specifications and efficiency of 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. Because this paper focuses on income per capita as the welfare aggregate, and because household surveys may not consistently report any additional variables, income per capita is used as the sole explanatory variable. This choice is not thought to affect results systematically or significantly. On the other hand, the choice of income over expenditure may affect the result systematically. Belhaj Hassine (2011) found that inequality in wage earnings in Egypt is nearly twice as high as inequality in expenditures. Hlasny and Verme (2013) found that the Gini of income per capita is systematically higher than the Gini for expenditure per capita by 2.5 percentage points.

Replacing

Another body of literature argues that the best approach to correct for poorly reported top incomes is to remove the top end of the distribution altogether and replace it with a parametric distribution. Atkinson et al. (2011), summarize literature contending that the distribution of top incomes is best illustrated 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 (2014) find that replacing actual top incomes with Pareto parametric estimates has a small but significant effect on the computed Gini. 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. Tested on Egypt, this approach yielded higher estimates than those reported by Hlasny and Verme (2014) for the same country. 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 non-response and its correction using Korinek et al.'s (2007) method. Second, we use the parametric properties of the Pareto distribution to evaluate how representative are the top income observations in our sample to the underlying income distribution. Third, following Cowell and Flachaire (2007) and Davidson and Flachaire (2007) we correct the Gini coefficient for the potential influence of top observations by replacing 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. Finally we compare the results with non-corrected Ginis or Ginis corrected for other statistical issues. This allows us to comment on the relative influence of extreme observations and other statistical issues in our data.

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 as follows:

$$F(x) = 1 - \frac{1}{x^{\alpha}}\ ,\ \ 1 \le x \le \infty\ , \tag{4}$$

where α is a fixed parameter called the Pareto coefficient and x is the variable of interest, in our case income per capita. It follows that the probability density function can be described as

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

The probability density function has the properties of being decreasing, tending to zero as x tends to infinity and with a mode equal to 1. Intuitively, 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.

The Gini coefficient under the estimated Pareto distribution for the k top-income households can be derived from the expression for the corresponding Lorenz curve (expression inside of the integral below) as

$$Gini = 1 - 2\int_0^1 1 - [1 - F(x)]^{1-1/\alpha}\, dF(x) = \frac{1}{2\alpha - 1} \tag{6}$$

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 $4\alpha(\alpha-1)/[n(\alpha-2)(2\alpha-1)^2(3\alpha-2)]$ (Modarres and Gastwirth 2006), and estimation error due to imprecision in the estimation of α is equal to $\eta/(2\alpha^2 - 2\alpha - 2\alpha\eta + \eta + 0.5)$, where η 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

$$Gini_{semi} = (1 + Gini_k)\frac{k}{n}s_k - (1 - Gini_{n-k})\left(1 - \frac{k}{n}\right)(1 - s_k) + \left(1 - \frac{2k}{n}\right). \tag{7}$$

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 parameter 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, and that the inequality index is sensitive to this. A 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. A higher semi-parametric Gini would indicate that the observed top incomes are lower than what the Pareto distribution would predict, potentially implying under-representation or measurement errors in relation to high-income units in the sample.

Modalities of the method

Estimation of parameters. One possible definition of the Pareto coefficient (α) as well as the inverted Pareto coefficient (β) as proposed in Atkinson et al. (2011) is:

$$\alpha = \frac{1}{1 - \left[\log\left(\frac{s10}{s1}\right) / \log(10)\right]} \tag{8}$$

$$\beta = \frac{\alpha}{\alpha - 1}, \tag{9}$$

where $s10$ and $s1$ represent the income shares of the top 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.

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. In what follows, we will report both coefficients but, 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

$$\alpha = \frac{1}{k^{-1}\sum_{i=0}^{k-1}\log X_{(n-i)} - \log X_{(n-k+1)}}, \tag{10}$$

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. In this paper, we estimate α using maximum-likelihood methods to obtain an estimate with a robust standard error. All these alternative estimation methods allow weighting of observations by their sampling probability, and yield similar results.

Appropriate parametric distribution. While Pareto distribution approximates 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 entire income distributions (McDonald, 1984). The upper tail of the distribution is heavy and decays like a power function. Four estimable parameters give the distribution flexibility to fit various empirical income distributions. The cumulative distribution function of the GB2 distribution is

$$F(x) = I\left(p, q, \frac{(x/b)^a}{1+(x/b)^a}\right) \tag{11}$$

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 one (McDonald, 1984).

The Gini index of income inequality under the GB2 distribution can be computed by evaluating the generalized hypergeometric function ${}_3F_2$ with the estimated parameters as arguments, and its standard error can be computed using the delta method (McDonald, 1984; Jenkins, 2009). In this paper, fit of the survey data to both the Pareto and the GB2 distributions will be evaluated.

Selection of parametric values for replacement of unreliable incomes. 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.[3] 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:

$$\widehat{Gini}_q = \sum_{i=1}^{m} Gini_{qi} / m \tag{12}$$

[3] Since top incomes in the US CPS do not appear to follow Pareto distribution exactly, Jenkins et al. fit the GB2 distribution instead. They replace top-coded values with random draws from the estimated GB2 distribution. 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.

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:

$$\widehat{var} = \frac{\sum_{i=1}^{m} \left(Gini_{q\,i} - \widehat{Gini_q}\right)^2 \Big/ (m-1)}{m} + \sum_{i=1}^{m} var_{qi} \Big/ m. \tag{13}$$

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 or the GB2 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).

3. Data

The methodologies outlined in the above section are evaluated and compared using three household surveys with vastly different characteristics: 1) the 2009 and 2011 rounds of the EU Statistics on Income and Living Conditions survey; 2) various rounds of the Current Population Survey, March Annual Social and Economic Supplement with the 2013 round serving as the primary round for our analysis, and 3) the 2009 Egyptian Household Income, Expenditure and Consumption Survey. These surveys can be viewed as prototypes of surveys with different types of problems related to measurement issues that affect top incomes and inequality estimates.

The EU Statistics on Income and Living Conditions (EU-SILC) survey, coordinated by a Directorate-General of the European Commission, Eurostat, covers one of the most heterogeneous and largest common markets, including some of the world's most affluent nations as well as former socialist economies. All European Union member states as well as Iceland, Norway and Switzerland are included. Incomes in the EU-SILC survey exhibit substantial cross-country inequality, but relatively less inequality within countries, as evidenced by the difference between state-specific and EU-wide Gini indexes (Table 1). The data include relatively large sample sizes for each state but suffer from very different non-response rates across member states and limited regional disaggregation. Non-response rates in the 2011 EU-SILC survey range from 3.3 to 50.7 percent across member states (3.5 to 48.1 percent in 2009). These features allow for a limited number of methods to be used to reevaluate inequality under various measurement issues.[4]

EU-SILC data are rarely used as one dataset for cross-country analysis in the same fashion as one would do cross-region analysis in a specific country. That is because EU-SILC data are derived from country specific surveys which may take different forms in different countries. However, in our case, they are an

[4] For more information on the EU-SILC see: http://ec.europa.eu/eurostat.

interesting set of data in that they are characterized by extreme diversity. They are therefore a good benchmark to test how different top incomes correction methodologies perform under such diversity.[5]

In what follows, we will make use of the newest round of the EU-SILC, that is, the 2011 round, and we will report on the 2009 round only on the margins. When not noted explicitly, the discussion refers to the 2011 round. Table 1 presents a summary of the 2011 EU-SILC data. (Table A1 in the annex presents a summary of the 2009 data.)

[Table 1]

The US Current Population Survey, March Annual Social and Economic Supplement (US-CPS) covers one of the most affluent countries, but the population it covers is relatively homogeneous between states. Incomes in the CPS exhibit a high degree of income inequality, particularly within US states rather than across states (Table 2). The CPS provides a large regionally well-disaggregated sample, but still suffers from a high rate of unit non-response of 9.5 percent nationwide in year 2013, ranging from 4.1 to 15.3 percent across individual US states.[6] Refer to table 2.

One problem with this survey is that the various components of income are top-coded. The technique used for top-coding is "rank proximity swapping," whereas 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. In addition, 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 the CPS (2013, Chart 1). 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).[7] To explore how influential this survey feature is, we could flag households with some of their income top-coded, and we could measure sensitivity of the measure of inequality to adjustments in their overall incomes. However, because this issue is absent in the EU-SILC and the Egyptian HIECS and because such flags are rare in survey data worldwide, we do not take advantage of the household-level flags in the CPS data in this paper.

[Table 2]

The 2009 Egyptian Household Income, Expenditure and Consumption Survey (EG-HIECS) is taken as an example of survey administered in an emerging or developing economy. Surveys in these countries are

[5] Indeed, sampling weights in the EU-SILC are distributed very widely, from essentially zero to 38,357.27 (mean 719.59, standard deviation 1,088.41) in the 2011 round. This compares to weights of 90.12 to 548.06 in the Egyptian HIECS (mean 370.65, st. dev. 59.54), and weights of 98.86 to 8,761.64 in the 2013 round of the US CPS (mean 1,904.67, st. dev. 971.66). This also suggests that comparing unweighted, EU-SILC weighted, and our non-response probability weighted statistics may yield very different estimates, much more so than in the US CPS or the Egyptian HIECS. Moreover, sampling weights in the EU-SILC are trimmed from below and from above to limit the extent to which individual observations can influence sample-wide statistics. To evaluate how much this trimming affects survey-wide results, we could compare results across alternative weighting schemes, or replace the trimmed weights with imputed values.

[6] 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 non-response, following Korinek et al. (2007).

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

characterized by reduced non-response rates as compared to wealthy countries while the statistical agencies that administer these surveys tend to refrain from applying post-survey censoring or data modifications. The Central Agency for Public Mobilization and Statistics (CAPMAS), the agency that administers the HIECS, has expended significant resources to ensure data completeness and reliability, as summary statistics show (Table 3). The CAPMAS does not apply data modification methods such as top coding, imputation of values or trimming of sampling weights. Item non-response is not an important issue in HIECS and unit non-response for the 2009 survey was about 3.7 percent, an extremely low value if compared to wealthy countries. However, unit non-response was systematic and influential to the measurement of inequality and the reason for non-response was not known (Hlasny and Verme 2013).[8] As shown in table 3, inequality within as well as across governorates is moderate, as the governorate-level and overall Gini coefficients indicate.[9]

[Table 3]

4. Results

Recall that we want to correct the Gini measure of inequality for top income biases and that, in doing so, we focus on two classes of methods. *Reweighting* methods initially designed to address top income biases generated by unit non-responses and *replacing* methods initially designed to address top income biases generated by outliers or artificially modified distributions (trimming, top-coding, etc). Results are presented following this classification. Note that *reweighting* can be used to address issues like trimming and top-coding and, vice-versa, *replacing* can be used to address issues of unit non-response. The two methods fundamentally address the same problem of top incomes biases but they were initially motivated by the different issues described.

Reweighting

Table 4 presents the benchmark results of this study, correcting distribution of incomes in the three household surveys for unit non-response using cross-state information. Following the lead of Korinek et al.'s (2006, 2007) and Hlasny and Verme's (2014) studies, these models estimate survey-response probability as a logistic function of the logarithm of income per capita. Logarithmic specification allows the relationship between income and response probability to be highly nonlinear, with the response rate dropping rapidly in the highest range of incomes. $g(x)$ in equation 1 is thus a parsimonious logarithmic function of a single variable: $g(income)=\theta_1+\theta_2 log(income)$, where income could also be represented by expenditure, consumption or other demographic variable deemed relevant, depending on data availability. This specification is thought to be robust and quite efficient in the measure of fit achieved. Since the explanatory variable (income, expenditure or consumption per capita) is available in all budget surveys while other demographic information may not be, this specification is also preferable as most useful to practitioners. In what follows, we will use income per capita. Income is the welfare variable that is most

[8] Jolliffe et al. (2004) explain why the distribution of consumption data in the HIECS may not be comparable to those in other surveys, essentially due to the way of accounting for values of durable goods.
[9] For more information on the Egyptian HIECS see www.capmas.gov.eg.

likely to be affected by measurement errors and top coding and the per capita form is chosen because income in household surveys is typically measured at the household rather than individual level.

The main finding in table 4 is that households' survey response probability is related negatively to income per capita. The coefficients on income $E(\theta_2)$ are consistently negative and highly significant, an indication that unit non-response 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. Initially ignoring sampling weights, and reweighting households by the inverse of their estimated response probability allows us to correct measures of inequality for the differential probability of rich and poor households to respond to the survey. Across the three household surveys, the corrected Gini coefficients are 38.70, 49.63 and 41.16. These are higher than the uncorrected and unweighted Gini coefficients by 0.21, 3.60 and 5.34 percentage points, statistically highly significant for the latter two.[10]

Making use of sampling weights provided by the national statistical agencies does not affect these findings. Applying the sampling weights to the distribution of incomes uncorrected for unit non-response leaves the Gini unchanged in the CPS and HIECS, and actually reduces the Gini in the EU-SILC by 5.9 percentage points. This is surprising, given that both correction schemes – correction for various sampling issues (including non-response in the case of the CPS and the EU-SILC), and correction for unit non-response were expected to inflate representation of atypical units such as top-income households. However, our correction for unit non-response significantly increases the estimate of inequality compared to both the unweighted and the sampling-weights corrected distributions of income.Applying both correction schemes in tandem, which is appropriate in the HIECS but amounts to double-correction for unit non-response in the CPS and the EU-SILC, leaves the basic findings above unchanged. The correction for unit non-response then amounts to 0.47, 3.86 and 4.79 percentage points of the Ginis across the three surveys, respectively.

[Table 4]

Given the significant correction for unit non-response identified in table 4, and the difference in the correction across the three household surveys, we should evaluate the implicit assumptions behind our model, as well as differences across the three household surveys.

Non-response rates. The results in table 4 suggest that the correction for unit non-response bias varies significantly across household surveys. Differences in non-response rates across the surveys do not explain the differences in the estimated bias satisfactorily. While the EU-SILC has the highest non-response rates, it is estimated to suffer from the lowest non-response bias in its Gini index.

Using a single survey for multiple years, we can evaluate how the varying unit non-response rates, and potentially also the changing extent or nature of inequality, affect the estimated bias. The US CPS is ideal for this exercise as it has been collected systematically for over fifty years, in a consistent format since 1989. Income distribution in the US CPS has also been consistent across years, with a moderate steady

[10] Table 5 reports that the correction varies from 0.35 to 9.66 percentage points across different waves of the US CPS. The small correction in the EU-SILC data for 2011 is consistent with that in the 2009 round. In the complete dataset of 30 member-states in the 2009 round of the EU-SILC survey, the sampling-weight uncorrected Gini coefficient is 43.30, while the one corrected for unit non-response is 43.42. The correction for unit non-response is 0.12 percentage points.

drift in mean incomes and the Gini coefficient. Using years 1989 through 2013, table 5 reports cross-sectional nationwide statistics and Gini coefficients.

[Table 5]

This analysis reveals that, even within a single survey administered across years, there is substantial variation in non-response rates, estimates of inequality, and estimates of the unit non-response bias. The estimated bias varies from 0.35% to 9.66% for the non-CPS weighted sample and from 0.35 to 10.23 for the CPS weighted sample. It depends positively on non-response rate, mean real income and corrected estimates of the Gini index (Pearson correlation of 0.63, 0.48 and 0.92, respectively, all statistically significant). Since unit non-response rates, real incomes and Gini index of inequality as measured in the US CPS have been persistently rising over time, the bias due to unit non-response has tended to increase. We also observe periods after 2005 when both mean income and the bias decrease. When these variables are studied jointly in a multiple regression, the bias turns out to be significantly affected by the true Gini index or the unit non-response rate, which are highly positively collinear. However, these facts still fall short of explaining credibly why the non-response bias estimated in the US CPS is significantly higher than that in the EU-SILC and slightly lower than that in the Egyptian HIECS.

Regional disaggregation. Next, we evaluate the role of the definition of regions j across which distributions of incomes are compared, and which dictate the size and number of errors to be minimized in the estimation of equation 2. Unfortunately, the three household surveys do not provide unit non-response rates for all administrative areas. The CPS includes information on Metropolitan-Core Based Statistical Areas (MCBSA) for approximately 75 percent of sampled households, at a similar rate for responding and non-responding households. This availability varies greatly across states (Table 2). In this section, we use a subsample from the 2013 US CPS for 24 states, each with MCBSA information available for over 75% of sampled households.[11]

With the HIECS, we face a similar problem. Greater level of disaggregation is available only for a 25% extraction from the HIECS sample rather than for the entire 100% sample. The CAPMAS provided the authors short-term access to the full 2009 HIECS on site in Cairo in May 2013. During the visit we performed the analysis at the governorate by urban–rural regions level (J=50 regions with N=939.7 responding households on average) and at the PSU level (J=2,526 with average N=18.6), but not at intermediate geographic levels. We report the analysis performed on the 25% extraction at the level of governorates (J=27 with average N=430.9), governorate by urban–rural regions (J=50 with average N=211.5), Kisms (J=446 with average N=26.1), groups of 1-32 Shakias within the same Kism (J=561 with average N=20.7), and PSUs (J=2,515 with average N=4.6).

Table 6 reports on the analysis performed at alternative degrees of geographic disaggregation for the 24-state subsample from the 2013 US CPS and for the 25% extraction from the 2009 HIECS. The table shows that the more detailed the degree of geographic disaggregation, the smaller the estimated bias due to unit non-response. In the CPS data uncorrected using sampling weights, the bias estimated using state-level disaggregation is 2.71 percentage points, falling to 1.58 points when estimated using MCBSA -level

[11] The CPS also includes information on counties, but only for 43% of households, and so this information cannot be effectively used.

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 (column 1), we obtain the same results as when we restrict our attention to households reporting their MCBSA (column 2), suggesting that the selection is not systematically related to income distribution or the tendency to respond to the survey. Similarly, in the Egyptian HIECS non-corrected by CAPMAS sampling weights, the analysis performed at greater degrees of disaggregation yields systematically lower estimates of the bias, from 4.24 to 3.38 percentage points.[12] The only exception to the consistent trend occurs in the case of disaggregation at the level of governorate urban-rural areas. While CAPMAS stratification methods account for differential sampling rates across urban and rural areas, it is possible that residents of these respective areas differ systematically in their behavioral responses, violating the assumption of stability across regions and confounding the results slightly. This could be attenuated by adding an urban-area indicator as an explanatory variable.

[Table 6]

Analyses using finer degrees of disaggregation yield lower corrections for unit non-response 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 non-response 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 non-response rates across regions.

The change in the estimated bias across different disaggregation methods is notably large for the US, where substantial income inequality exists at the sub-state level, across cities rather than across states, and where non-response rates are similar across states as well as across MCBSAs. The estimated bias varies much less across different disaggregation methods in Egypt, where inequality occurs across governorates with relatively less inequality within them, and non-response rates vary greatly across kisms or across more finely delineated regions.

In the CPS data, disaggregation from the state to the MCBSA level (eight times smaller regions) reduces the estimated bias to the Gini coefficient from 2.71 to 1.58 percentage points, by 42 percent. MCBSAs have non-response rates of 0.0–23.5%, or twice the cross-state range of non-response rates, 4.1–15.4%. In the HIECS data, disaggregation of a similar magnitude from governorate urban–rural strata to kisms

[12] A similar analysis performed on twenty US states, each with over 80% of households with known MCBSAs was also performed, with essentially the same results as in table 6. Also, a similar analysis performed on the full 100% sample of the HIECS at the governorate or PSU levels showed the same qualitative result – the smaller and more numerous the regions, the lower the estimated bias (Hlasny and Verme 2013).

reduces the estimated bias from 4.45 to 3.87 percentage points, by only 13 percent. Kisms have non-response rates of 0.0–30.0%, or three times the range of non-response rates across governorate urban–rural strata, 0.0–10.5%.

Optimal disaggregation. Since the degree of geographic disaggregation of survey sample affects the correction for unit non-response 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 can conduct a simple experiment. We first choose a high quality sample with low non-response rates 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. 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 use the 2010 sample of the US CPS, and the 25% sample of the 2009 Egyptian HIECS. The 2010 CPS sample covers 75,277 responding households with incomes per capita greater or equal to one, one of the largest samples across years. 73.5 percent of responding households and 79.2 percent of non-responding households have a known MCBSA. The unit non-response rate (7.01%) was one of the lowest across all the evaluated years, and the corresponding Gini bias estimated using state-level disaggregation, a mere 2.04 percentage points, was also among the lowest (table 5). This sample is as close to one free of unit-non-response problems as we can get. We can evaluate the correction for unit non-response bias performed at the level of Census regions, states and MCBSAs.The 25% sample of the 2009 HIECS also has a very low rate of household non-responses (3.71%), and a low estimated bias due to them (3.59 percentage points using governorate urban–rural area disaggregation). This sample can be disaggregated geographically by governorate, governorate urban–rural area, kism, or group of nearby shakias, with each region containing a sufficient number of responding households.[13]

For the trimming of observations, we apply the stochastic behavioral response proposed in equation 1 and confirmed in tables 4 and 6. Richer households have a lower propensity to appear in the sample.

[13] Analysis at the level of individual shakias or even PSUs is deemed not to be appropriate, since these regions cover as few as 3-5 responding households. The 100% sample of the HIECS would be necessary to conduct analysis disaggregated at this level successfully, but we currently do not have that dataset at our disposal.

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 for the US CPS and the Egyptian HIECS in table 4 (and similar to estimates across all columns of table 6), *g(x)* in equation 1 is taken to be: $g(income)=\theta_1+\theta_2 log(income)$=13.0-1.0*log(income)* for samples from both surveys.

Table 7 reports the results of this experiment. Across columns, different degrees of geographic disaggregation are evaluated. Across rows, different fractions of observations are trimmed from the sample according to the stochastic weighting scheme, with richer households systematically more likely to be trimmed. The top rows report on an experiment where 6.5% of observations were trimmed as non-responders.[14] The following rows trim 7%, 10%, 13% and 16% of observations. For the US CPS, as observations are trimmed, Gini in the sample falls from 46.54 in the original sample to 45.54 in the subsample with 16% of observations trimmed. While we would expect the Gini to keep falling as more of top incomes are trimmed, this does not occur consistently here due to the stochastic way in which observations were trimmed and due to the limited number of random draws (30) there were evaluated. For the Egyptian HIECS, the Gini falls nearly consistently from 36.57 in the original sample to 34.51 in the subsample with 16% of observations trimmed.

In the US CPS, our method correcting for unit non-response performs well when only 6.5–7% of the sample is non-responding, but the correction is too small when 10–16% are non-responders. Across all weighting schemes, geographic-aggregation methods and degrees of sample trimming, the correction slightly underestimates the unit non-response bias, since the corrected Ginis are all smaller than the true Gini. The corrections range between 0.1 and 1.04 percentage points of the Gini, and bridge between a tenth and nine-tenths (two-fifths on average) of the bias induced by unit non-response.

The method using state-level data aggregation performs consistently better than ones at the MCBSA or the Census-division levels. Clearly, seven Census divisions is too few to perform the fitting adequately. On the other hand, using 22 states or 171 metropolitan areas produces similar results. The under-correction is also statistically insignificant for the cases when only 6.5–7% of the sample is trimmed in the state-level or MCBSA-level analysis. Finally, comparing the state-level and MCBSA-level analysis, we find the expected result that the finer the degree of disaggregation – MCBSA rather than state level – the smaller the correction for unit non-response, corroborating the results in table 6. In this case, the reduction in the correction is damaging as it keeps the corrected Gini from reaching up to the true value.

Regarding derivation of the actual behavioral response function, the models in table 7 perform decently when 6.5–7% of the sample is trimmed, in models using state-level or MCBSA-level disaggregation. Estimated coefficients are within one standard deviation from the actual values (θ_1=13, θ_2=-1). When more observations are trimmed, or when Census division disaggregation is used, estimates differ from the actual values more, suggesting poor fit.

For the Egyptian HIECS, similar patterns emerge, although the results are unstable. For the most part, the corrections for unit non-response under-correct for the non-response bias, as all the corrected Ginis are

[14] The algorithm performing randomized trimming according to household weights could not trim fewer than 6.5% of observations in the US CPS sample, and fewer than 1.1% of observations in the 25% HIECS sample, while observing the desired weighting scheme.

lower than the true original-sample Gini. In any case, the corrections amount to 0.9–1.25 percentage points of the Gini, and bridge between a half and four-fifths (three-fifths on average) of the bias induced by unit non-response. The strength of correction as fraction of the bias falls when more observations are trimmed, confirming a finding from the US CPS.

Comparing the five columns for HIECS data, we find that the finer the degree of disaggregation – from governorates to PSUs – the smaller the correction for unit non-response tends to be. Like in the US CPS, this works against the goal of reaching up to the true statistic. Compared to the US CPS (columns 2–3 in table 7), however, the fall in the correction for non-response across columns is quite tepid in the HIECS (columns 4–8). This confirms the finding in table 6 that the comparison of demographic and behavioral factors across regions affects the relative performance of alternative ways of geographic disaggregation.

Like in the US CPS, the trend of bias corrections falling across columns is not strictly monotonic (particularly in the case of income distributions corrected by CAPMAS sampling weights), reflecting problems including 1) insufficient number of regions in the case of governorate level disaggregation; 2) limited comparability of households' behavioral responses across regions, particularly in less disaggregated samples; and 3) insufficient number of income observations in the case of PSUs. Geographic disaggregation that appears to provide the most consistent correction for non-response – both across unweighted and sampling-weights corrected income distributions, and across different degrees of trimming – is at the level of groups of nearby shakias.

In conclusion, table 7 provides several important insights regarding the performance of the method for correcting for the unit non-response bias through reweighting of income observations. The method performs best in samples with low or moderate non-response rates, while it appears to be imprecise in samples with high non-response rates. Analysis performed at an intermediate degree of geographic disaggregation yields better correction than disaggregation into too many or too few regions. Among the options considered for the US CPS, state-level disaggregation was clearly preferred, while for the 25% extraction of the Egyptian HIECS, the jury is out on disaggregation into kisms versus into groups of nearby shakias. In the 100% sample of the HIECS, groups of shakias or a similar degree of regional disaggregation would presumably be justified as the preferred method. With an arbitrary worldwide household survey, properties of the data at hand should guide the choice over the appropriate degree of disaggregation, and should guide our interpretation of estimates.

Replacing

We use a methodology first proposed by Cowell and Victoria-Feser (2007) to test sensitivity of the Gini coefficients to extreme observations on the right-hand side of the distribution. If top incomes turn out to be influential, in the raw income distribution as well as in the distribution corrected for unit non-response bias, we correct for their presence using an estimated Pareto distribution as discussed in the methodological part.

[Table 7]

Table 8 presents semi-parametric estimates of Gini coefficients, obtained by replacing the highest top 0.1–1.0 percent of income observations with values imputed from the corresponding Pareto distribution as per Cowell and Flachaire (2007), and Davidson and Flachaire (2007).[15] The first four rows show the benchmark non-parametric estimates from table 4 – unweighted; corrected for sampling probability using statistical-agency weights; corrected for non-response bias as per table 4; and corrected for both. The next four rows present the main results – semi-parametric estimates with the top 0.1 percent of incomes imputed from corresponding Pareto distributions. The four rows differ in the definition of the top 0.1 percent of incomes and in the estimated α, as they assign different weights to each top income observation (i.e., unity, sampling weights, non-response correcting weights, or both). The following twelve rows report on an analogous exercise, where the parametric imputation is performed on top 0.2, 0.5 or 1.0 percent of incomes.

[Table 8]

Table 8 shows that the exact cutoff for incomes to be replaced and the way income observations are weighted affect greatly the estimated shape of the top income distribution. For the EU-SILC, the estimated Pareto coefficient α varies between 1.65–2.09 and 2.36–3.12 depending whether only top 0.1% or up to top 1.0% of households are used for estimation. These ranges are 3.29–22.74 and 1.70–2.29 in the US CPS, and 0.81–2.07 and 1.75–2.51 in the Egyptian HIECS. The widths of these intervals also indicate that the estimated α depends on the way income observations are weighted. Most notably, the Pareto coefficients change systematically as more of top incomes in a distribution are evaluated.

In the EU-SILC and the Egyptian HIECS, the higher the fraction of incomes evaluated, the higher the Pareto coefficient (and the lower the corresponding inverted Pareto coefficient), and thus the lower the estimated top income share. That suggests that in the EU-SILC and the Egyptian HIECS extreme incomes may be a problem among the top-most 0.1% of incomes, but not as much among the following 1% of incomes. In the US CPS, the opposite phenomenon occurs: income share of the handful super-rich (top 0.1%) households is estimated to be not as high as in other income distributions or under a smooth Pareto curve, but income share of the next 1% of incomes is higher. One likely reason of this finding is that top-most incomes in the CPS data are top-coded via ‘rank-proximity swapping’ and rounding.

The estimated Gini coefficients are affected by the method of modeling top incomes in a qualitatively similar fashion, but to a much lower degree. The correction for potentially extreme or imprecise top income observations results in a reduction of up to 0.005 percentage points in the EU-SILC and 0.014 percentage points in the HIECS, and an increase of up to 0.019 percentage points in the CPS. Half of the Gini corrections across the three surveys are downward and half are upward, and the corrections grow in absolute value with the fraction of observations replaced, but are all trivial.[16,17] It appears that the exact

[15] Table A3 in the annex shows the analogous results for the exercise replacing the highest top 5%, 10% or 20% of income observations with values under the Pareto distribution. These high percentages of top incomes are chosen to allow precise estimation of Pareto coefficients. It is also 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). Table A3 is comparable to Hlasny and Verme’s (2013) table 3. The results in table A3 are more stable than in table 8, because a larger fraction of incomes, and thus even values not too extreme are being replaced.

[16] In table A3, the corrections are larger, because greater fractions of observations are replaced with fitted values. The correction is up to 0.24 percentage points in absolute value in the EU-SILC (from 44.10 to 44.35), up to 0.25 percentage points in the CPS (from 46.16 to 46.41), and up to 0.56 percentage points in the HIECS (from 41.16 to

values of top-most incomes are not influential for the measurement of inequality in the overall income distribution, as compared to the corresponding smooth Pareto dispersion of top incomes, because they may skew Gini estimates only slightly upward or downward. In perspective of the findings in preceding sections we conclude that the systematic under-representation of top income households due to unit non-response is a far more worrisome problem biasing inequality estimates systematically downward.

*Parameter specifications***.** One potential criticism of the above approach is that it relied on the fit of true top incomes to the one-parameter Pareto distribution. While the Pareto distribution has been accepted as providing a good fit for many national income distributions around the world, its fit to the CPS data has been questioned. Several studies have suggested other, more flexible statistical distributions as providing a better fit, such as the three-parameter Singh-Maddala and Dagum distributions. These are limit cases of a four-parameter generalized beta (type 2) distribution. In this section we re-estimate the semi-parametric Gini coefficients assuming top incomes to be distributed as under the generalized beta distribution.

Table 9 reports the results.[18] Coefficient estimates in table 9 carry small standard errors and are quite consistent across different weighting schemes of the samples, particularly for the US CPS and the Egyptian HIECS. For the EU-SILC, the coefficients – as well as the inferred parametric and semiparametric Ginis – vary across columns, due to heterogeneity across member-states and great differences in the alternative weights imposed. The coefficient estimates imply that the generalized beta distribution cannot be easily approximated by Singh-Maddala or Dagum distributions because *E(p)* and *E(q)*, respectively, are significantly different from unity across all surveys and most columns. Only in three columns, all using corrections for unit non-response, there is some support for one of these two alternative distributions, as the estimate of *E(p)* in column 3 and the estimates of *E(q)* in columns 7 and 8 are within two standard errors of unity.

Comparing the Ginis in table 9 to the nonparametric estimates in table 4, we find that the parametric and semi-parametric Ginis under the assumed generalized beta distribution tend to be lower, implying that the true incomes are distributed more unequally than incomes predicted under that distribution. This is

---

40.60). Greater corrections in absolute value occur when a greater number of top income observations are replaced – the corrections are greatest when top 20% of income observations are replaced. The corrections to the Gini tend to be positive in the EU-SILC and the CPS, suggesting that actual incomes there are lower or distributed more narrowly than would be predicted under the corresponding Pareto distributions. The corrections to the Gini are overall negative in the HIECS, suggesting that incomes observed there are higher or distributed more widely than would be predicted under the corresponding Pareto distributions.

[17] A final note is that the parametric estimates of the Gini among top incomes in table 8 were calculated under smooth fitted Pareto curves rather than from any observations or fitted values per se. As a robustness check, we have re-estimated these Ginis by replacing top incomes with randomly drawn numbers from the corresponding Pareto distributions, then repeating the exercise 30 times and taking an average of the 30 obtained Ginis (refer to equation 12). These Ginis from random draws differ by -1.28 to +1.53 percentage points from the smooth-distribution Ginis in table 8 (mean difference +0.02, mean difference in absolute value 0.50). Still, the corrections of the nonparametric Gini coefficients are very similar to those obtained in table 8.

[18] An estimation note is in order: During estimation on the HIECS with the CAPMAS-provided sampling weights the algorithm fitting a generalized beta distribution had trouble converging due to the bottom one income observation (450 Egyptian pounds/year). Similarly, during estimation on the EU-SILC with the survey-provided sampling weights and non-response weights, the algorithm had trouble converging due to the bottom two income observations (2.43–2.50 Euro/year). These estimation issues indicate atypical distribution of the bottom-most incomes in the two surveys. Indeed, there are over 100 observations in the EU-SILC with annual income less than 100 Euro, suggesting measurement errors.

particularly true for the HIECS, where the downward correction of the Gini is up to 3 percentage points and typically 1.5 percentage points, and less so for the EU-SILC (correction of up to 1.1 and typically 0.4 percentage points) and for the CPS (correction of up to 0.6 and typically 0.2 percentage points). Using random income draws from a generalized beta distribution produces a similar correction of the Gini as numerical inference of the Gini under a smooth distribution, verifying that the procedure works correctly.

Compared to the Pareto distribution evaluated in the previous section, the corrections to the Gini coefficients under the generalized beta distribution are larger and consistently negative for all three surveys.[19] This indicates that the estimated generalized beta distributions predict a narrower dispersion of top incomes than the estimated Pareto distributions. For the EU-SILC and the Egyptian HIECS, the downward correction to the Gini derived in the previous section is now estimated to be even larger, of up to 1.1 percentage points for the EU-SILC and up to 2.9 percentage points for the HIECS. For the US CPS, the small upward correction to the Gini derived in the previous section is now replaced by a small downward correction, of up to 0.8 percentage points. This suggests that our assumption about the distribution of true top incomes affects our correction for extreme observations. In absolute terms, however, the difference is modest, at 0.1–1.1 percentage points (mean 0.5) for the EU-SILC, 0.0–0.8 percentage points (mean 0.3) for the CPS, and 0.0–3.0 percentage points (mean 1.2) for the HIECS.

[19] Because top-income Gini coefficients are derived 'quasi-nonparametrically' and averaged across 30 random draws from the smooth distribution, there are 14 instances out of 96 where the generalized-beta Gini is higher than the semi-parametric Pareto Gini (tables 8 and A3).

Table 1. Non-response rate and income distribution by member state, 2011 EU-SILC

Member State	Households	Non-response Rate (%)	Mean Equivalised Disposable Income per Capita (Euro)	Member State Gini, EU-SILC weighted households
Austria	6,183	22.6	23,713.37	27.59
Belgium	5,897	36.7	21,622.14	27.63
Bulgaria	6,548	7.5	3,415.42	35.99
Cyprus	3,916	10.2	20,084.84	31.65
Czech Republic	8,865	17.1	8,402.77	25.91
Denmark	5,306	44.4	28,441.21	27.45
Estonia	4,980	26.0	6,475.47	32.62
Finland	9,342	18.1	23,870.09	27.28
France	11,348	18.0	24,027.78	30.84
Germany	13,473	12.6	21,496.55	30.21
Greece	5,969	26.5	12,704.72	32.92
Hungary	11,680	11.2	5,146.29	26.86
Iceland	3,008	24.8	20,668.26	24.99
Italy	19,234	25.0	18,353.37	31.72
Latvia	6,549	18.9	5,048.72	34.98
Lithuania	5,157	18.6	4,588.81	33.02
Luxembourg	5,442	43.3	37,232.63	27.32
Malta	4,070	11.8	12,167.55	28.29
Netherlands	10,469	14.5	22,726.06	25.66
Norway	4,621	50.7	38,616.14	24.98
Poland	12,861	14.9	5,849.61	32.10
Romania	7,614	3.3	2,447.42	32.37
Slovakia	5,200	14.5	6,983.48	26.21
Slovenia	9,246	23.8	12,714.07	25.84
Spain	12,900	37.2	14,584.40	32.67
Sweden	6,694	36.5	23,727.45	25.76
United Kingdom	8,009	27.3	20,843.59	32.85
Wtd. Mean (Total)	7,947 (214,581)	23.60	17,727.32	30.68 (38.23)

Note: Non-response rate is reported in the member-states' Intermediate/Final Quality Reports at the state level as *NRh* for total sample. Croatia, Ireland, Portugal and Switzerland did not submit their Quality Reports. Per-capita income is weighted by household size. Incomes less than 1 are omitted. Mean incomes may not be representative of those for the entire states, as they omit non-responding households. For clarity of presentation, Ginis are multiplied by 100.

Table 2. Non-response rate and income distribution by state, 2013 CPS March Supplement

State	Metrop. CBSAs	MCBSA Known (% hhds.)	Responding Households	Non-response Rate (%)	Mean Income per Capita ($)	Gini, CPS-wtd. hhds
Alabama	8	71.2	818	6.2	24,138.77	45.09
Alaska	0	0.0	859	12.7	29,041.85	43.01
Arizona	3	84.7	934	8.4	23,518.89	48.32
Arkansas	3	49.5	826	5.6	21,019.29	46.59
California	23	98.4	6,747	8.6	27,525.20	49.88
Colorado	6	89.0	1,646	9.2	30,117.24	43.89
Connecticut	6	92.1	1,592	12.5	36,135.82	44.85
Delaware	2	79.8	1,134	8.2	25,528.58	43.28
Distr. Columbia	1	100.0	1,297	13.3	45,482.45	50.79
Florida	19	96.2	3,136	5.1	25,703.22	44.75
Georgia	10	82.0	1,608	6.9	25,285.30	46.00
Hawaii	1	70.2	1,215	6.8	27,270.77	46.12
Idaho	2	46.7	767	8.9	22,251.83	44.96
Illinois	10	89.1	2,240	8.3	29,677.16	47.62
Indiana	9	69.7	1,091	8.3	24,372.35	42.85
Iowa	6	49.1	1,361	7.1	26,319.05	40.47
Kansas	4	65.1	1,049	8.8	26,200.10	43.19
Kentucky	4	48.0	1,031	8.2	22,601.75	39.90
Louisiana	6	82.9	754	7.4	22,305.71	43.22
Maine	2	40.3	1,172	13.4	26,789.77	41.26
Maryland	4	92.7	1,736	15.4	33,467.77	43.67
Massachusetts	6	94.1	1,070	12.9	31,864.75	44.69
Michigan	12	83.4	1,636	9.8	26,922.60	45.61
Minnesota	3	70.2	1,706	9.1	29,875.77	40.22
Mississippi	3	33.4	712	8.1	21,183.25	49.39
Missouri	5	70.5	1,151	8.4	26,928.65	43.80
Montana	1	12.5	707	4.1	24,531.35	40.25
Nebraska	1	40.8	1,104	9.3	26,174.53	39.29
Nevada	2	87.7	1,147	10.4	24,051.55	45.31
New Hampshire	2	41.9	1,402	12.5	32,411.46	40.32
New Jersey	7	100.0	1,412	13.5	33,882.08	45.11
New Mexico	4	69.7	726	7.6	28,928.90	53.08
New York	9	92.3	3,143	13.9	28,819.80	48.91
North Carolina	9	64.3	1,520	8.7	23,821.30	43.96
North Dakota	1	26.7	922	6.9	30,477.18	44.41
Ohio	9	75.7	1,961	10.2	24,904.71	42.32
Oklahoma	3	67.6	906	7.0	24,216.52	46.29
Oregon	5	76.4	1,012	11.8	25,489.51	42.70
Pennsylvania	11	82.6	2,197	9.6	27,146.49	44.15
Rhode Island	1	100.0	1,192	15.3	30,503.79	47.27
South Carolina	8	66.8	1,016	6.2	23,168.42	41.56
South Dakota	1	27.2	1,065	8.0	25,255.50	42.20
Tennessee	6	65.8	1,003	8.7	23,283.45	45.52
Texas	17	86.4	4,310	9.8	24,270.31	48.45
Utah	3	77.9	861	6.9	22,753.38	43.79
Vermont	1	32.4	964	14.8	28,701.23	41.51
Virginia	6	82.4	1,568	8.8	32,788.31	45.66
Washington	7	83.1	1,283	9.8	29,870.95	45.89
West Virginia	2	28.6	716	6.6	23,647.00	42.89
Wisconsin	10	69.6	1,405	6.7	27,626.79	41.37
Wyoming	0	0.0	935	9.7	27,221.34	43.16
Wtd. Mean (Total)	5.57 (284)	74.6	1,446 (73,765)	9.5	27,463.41	45.15 (46.16)

Notes: MCBSA availability is reported for both responding and non-responding households. Non-response 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 3. Non-response rate and income distribution by governorate, 2009 HIECS (100%)

Governorate	PSUs	Households	Non-response Rate (%)	Mean Income per Capita (E)	Governorate Gini, CAPMAS-Weighted Hhds.
Alexandria	149	2,801	6.0	5,393.10	32.57
Assiut	101	1,872	2.4	2,665.06	34.18
Aswan	52	978	1.0	3,635.79	29.67
Behera	152	2,871	0.6	3,680.44	25.00
Beni Suef	69	1,294	1.3	2,887.36	25.91
Cairo	285	5,194	8.9	6,499.94	40.69
Dakahlia	176	3,289	1.6	4,467.94	28.30
Damietta	52	959	2.9	5,460.37	27.45
Fayoum	78	1,466	1.1	3,071.68	25.56
Gharbia	139	2,584	2.2	4,606.58	30.13
Giza	215	3,939	6.5	4,347.80	38.44
Ismailia	52	967	2.1	5,401.84	40.66
Kafr ElSheikh	85	1,547	4.2	4,279.37	28.02
Kalyoubia	145	2,668	3.2	4,137.20	29.97
Luxor	14	263	1.1	4,704.10	31.56
Matrouh	11	209	0.0	5,861.38	37.12
Menia	128	2,371	2.5	3,451.37	31.49
Menoufia	107	1,977	2.8	4,147.15	31.06
New Valley	8	146	3.9	5,322.18	26.31
North Sinai	14	243	10.5	3,768.41	27.73
Port Said	50	925	7.4	6,501.37	35.84
Qena	88	1,628	2.6	3,302.03	28.66
Red Sea	13	239	3.2	7,050.69	38.47
Shrkia	175	3,262	1.9	3,662.45	27.60
South Sinai	4	69	9.2	10,969.95	68.00
Suez	50	951	4.9	7,269.37	32.68
Suhag	114	2,145	1.0	2,809.37	28.44
Wtd. Mean (Total)	94 (2,526)	1,735 (46,857)	3.7	4,653.03	31.76 (35.56)

Notes: Non-response rate, reported in the survey at the PSU level, is weighted by the number of responding households in each PSU. Per-capita income and expenditure are further weighted by household size. Mean incomes may not be representative of those for the entire governorates, as they omit non-responding households. For clarity, Ginis are multiplied by 100.

Table 4. Benchmark results of Gini correction for unit non-response bias

	EU-SILC (2011)	US CPS (2013)	HIECS (2009), 100% sample*
$E(\theta_1)$	6.998	12.959	12.948
(s.e.)	(2.302)	(2.444)	(0.070)
$E(\theta_2)$	-0.601	-1.032	-1.138
(s.e.)	(0.231)	(0.226)	(0.008)
Regions *j*	27 member states	51 states	55 governorate urban–rural areas
Households *i*	214,581	73,765	46,857
Uncorrected Gini	44.10	46.03	35.82
	(0.09)	(0.18)	(0.35)
Gini using stat. agency weights	38.23	46.16	35.56
	(0.14)	(0.24)	(0.32)
Gini corrected for unit non-response bias	**44.31**	**49.63**	**41.16**
	(0.23)	(0.44)	(2.04)
Gini corrected for unit non-resp. bias & with stat. agency weights	**38.70**	**50.02**	**40.35**
	(0.26)	(0.59)	(1.73)
Unit non-response bias	**0.21**	**3.60**	**5.34**
Bias (using stat. agency weights)	**0.47**	**3.86**	**4.79**

Notes: For clarity, Ginis and their standard errors are multiplied by 100. Standard errors on Ginis are bootstrapped. Only incomes greater or equal to 1 are retained. Note that results for the 2009 HIECS differ from those of Hlasny and Verme (2013) mainly because of the choice of the welfare aggregate and explanatory variable (income per capita in this paper and expenditure per capita in Hlasny and Verme, 2013).

Table 5. Gini correction for unit non-response bias across years 1989-2013, US CPS

Year	Mean income/ capita ($)	Real mean income/ capita (2000$)	Non-respons e rate (%)	Gini, non-weighted	Gini, CPS-wtd. hhds	Gini, non-response corrected	Gini, non-resp. & CPS wghted.	Unit non-response bias in Gini	Bias in CPS wghted .Gini
1989	12,589.96	17,483.79	5.46	**41.74 (.14)**	42.05 (.17)	**44.00 (.28)**	44.32 (.30)	**2.26**	**2.27**
1990	13,430.92	17,695.52	4.55	**42.10 (.13)**	42.10 (.15)	**44.74 (.29)**	44.83 (.32)	**2.64**	**2.73**
1991	13,694.75	17,314.51	4.70	**42.06 (.13)**	42.00 (.15)	**44.77 (.29)**	44.72 (.32)	**2.71**	**2.72**
1992	14,009.88	17,195.30	5.02	**42.01 (.13)**	41.95 (.15)	**43.52 (.22)**	43.47 (.22)	**1.50**	**1.52**
1993	14,337.65	17,086.11	4.99	**42.15 (.13)**	42.08 (.15)	**43.60 (.19)**	43.50 (.22)	**1.45**	**1.42**
1994	14,791.24	17,186.58	5.17	**42.51 (.13)**	42.41 (.15)	**45.73 (.36)**	45.30 (.32)	**3.23**	**2.89**
1995	15,304.90	17,293.33	4.53	**42.77 (.15)**	42.62 (.17)	**43.12 (.15)**	42.97 (.17)	**0.35**	**0.35**
1996	16,780.93	18,417.31	7.69	**44.59 (.19)**	44.41 (.21)	**49.48 (.52)**	49.37 (.58)	**4.89**	**4.96**
1997	17,648.09	18,934.59	7.18	**45.29 (.20)**	45.14 (.22)	**49.80 (.42)**	49.78 (.46)	**4.51**	**4.64**
1998	18,808.96	19,870.56	7.79	**45.54 (.20)**	45.33 (.21)	**52.60 (.55)**	52.37 (.58)	**7.05**	**7.04**
1999	19,722.67	20,385.61	7.90	**45.08 (.18)**	44.88 (.20)	**49.77 (.41)**	49.58 (.45)	**4.69**	**4.70**
2000	20,204.57	20,204.57	6.89	**44.35 (.16)**	44.33 (.18)	**48.56 (.32)**	48.67 (.37)	**4.20**	**4.34**
2001	21,517.55	20,922.20	8.03	**44.95 (.18)**	45.02 (.21)	**50.56 (.44)**	50.74 (.51)	**5.61**	**5.72**
2002	21,209.13	20,301.35	7.31	**44.99 (.17)**	45.50 (.22)	**48.62 (.31)**	49.34 (.39)	**3.63**	**3.84**
2003	21,227.65	19,866.31	7.17	**44.91 (.18)**	45.41 (.22)	**49.46 (.40)**	50.11 (.48)	**4.56**	**4.70**
2004	21,766.41	19,842.12	7.69	**44.78 (.16)**	45.27 (.21)	**50.58 (.46)**	51.42 (.57)	**5.80**	**6.15**
2005	22,642.33	19,964.20	9.01	**44.73 (.16)**	45.25 (.20)	**54.39 (.63)**	55.48 (.78)	**9.66**	**10.23**
2006	23,810.11	20,337.80	8.61	**45.20 (.16)**	45.64 (.20)	**53.73 (.56)**	54.99 (.71)	**8.53**	**9.35**
2007	25,122.74	20,868.96	8.66	**45.12 (.15)**	45.49 (.19)	**49.65 (.35)**	50.11 (.41)	**4.53**	**4.62**
2008	25,763.93	20,606.07	7.82	**44.68 (.14)**	44.99 (.17)	**47.78 (.28)**	47.99 (.31)	**3.09**	**3.00**
2009	26,059.47	20,916.86	7.06	**44.70 (.15)**	45.11 (.18)	**47.19 (.24)**	47.66 (.29)	**2.48**	**2.55**
2010	25,578.70	20,199.64	7.01	**45.24 (.15)**	45.48 (.18)	**47.28 (.22)**	47.62 (.28)	**2.04**	**2.14**
2011	25,683.59	19,661.84	8.12	**45.21 (.17)**	45.68 (.23)	**46.94 (.32)**	47.67 (.46)	**1.73**	**1.99**
2012	26,773.36	20,080.55	8.93	**45.71 (.18)**	46.20 (.24)	**49.21 (.42)**	50.17 (.60)	**3.50**	**3.97**
2013	27,463.41	20,300.74	9.54	**46.03 (.18)**	46.16 (.24)	**49.63 (.44)**	50.02 (.59)	**3.60**	**3.86**

Notes: Real incomes are computed using CPI with year 2000 as base. For clarity, Ginis and their standard errors are multiplied by 100. Standard errors on Ginis, in parentheses, are bootstrapped.

Table 6. Gini correction for unit non-response bias, varying geographic disaggregation

	2013 CPS: 24 states	2013 CPS: 24 states, households with known MCBSA		2009 HIECS, 25% sample				
	Analysis at the state level	state level	MCBSA level	governorate level	governorate urban–rural level	level of kisms	level of nearby shakias	PSU level
$E(\theta_1)$	12.101	11.704	9.925	10.895	10.964	10.587	10.575	10.301
(s.e.)	(3.511)	(3.736)	(2.319)	(0.080)	(0.070)	(0.020)	(0.018)	(0.012)
$E(\theta_2)$	-0.954	-0.918	-0.751	-0.904	-0.913	-0.872	-0.870	-0.839
(s.e.)	(0.325)	(0.346)	(0.218)	(0.008)	(0.008)	(0.002)	(0.002)	(0.001)
Regions *j*	24 states	24 states	185 MCBSAs	27 governt.	50 urban v. rural governt.	446 kisms	561 groups of nearby shakias	2,515 PSUs
Households *i*	45,616	40,746	40,746	11,634	11,634	11,634	11,634	11,634
Households per region	1,900.67	1,697.75	220.25	430.89	211.53	26.09	20.74	4.63
Uncorrected Gini	47.27 (0.23)	47.49 (0.25)		36.57 (0.96)				
Gini using stat. agency weights	46.87 (0.29)	47.02 (0.31)		36.01 (0.76)				
Gini corrected for unit non-response bias	**50.35** (0.50)	**50.21** (0.50)	**49.07** (0.39)	**40.81** (2.99)	**41.02** (3.10)	**40.44** (2.78)	**40.39** (2.76)	**39.95** (2.53)
Gini corrected for unit non-response with stat. agency wghts.	**50.25** (0.65)	**50.18** (0.65)	**48.91** (0.51)	**39.90** (2.69)	**39.60** (2.38)	**39.85** (2.38)	**39.81** (2.36)	**39.65** (2.17)
Unit non-response bias	**3.08**	**2.71**	**1.58**	**4.24**	**4.45**	**3.87**	**3.82**	**3.38**
Bias (using stat. agency weights)	**3.38**	**3.16**	**1.89**	**3.89**	**3.59**	**3.84**	**3.80**	**3.64**

Notes: For clarity, Ginis and their standard errors are multiplied by 100. Standard errors on Ginis are bootstrapped. 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 7. Gini correction for unit non-response bias in a trimmed sample

	2010 CPS: 22 states, households with known MCBSA (N=38,641)			2009 HIECS, 25% sample (N=11,634)				
True uncorrected Gini	46.544 (0.200)			36.568 (0.958)				
True Gini using stat. wghts	46.312 (0.239)			36.006 (0.761)				
Disaggregation into regions *j*	7 Census divisions	22 states	171 MCBSAs	27 governt.	55 governt. urban–rural	446 kisms	561 groups of nearby shakias	2,515 PSUs
	6.5% or 2,512 trimmed, N=36,129 **Uncorrected Gini: 45.449 (0.193)** **Gini using CPS wghts: 45.210 (0.230)**			**6.5% or 756 trimmed, N=10,878** **Uncorrected Gini: 34.797 (0.741)** **Gini using CAPMAS weights: 34.364 (0.616)**				
Gini corrected for unit non-response	45.690 (0.190)	46.474 (0.249)	46.343 (0.239)	35.930 (0.832)	35.865 (0.810)	35.826 (0.795)	35.833 (0.798)	35.757 (0.779)
Gini corrected for unit non-response & sampling wghts	45.466 (0.227)	46.265 (0.302)	46.130 (0.291)	35.525 (0.734)	35.409 (0.665)	35.546 (0.706)	35.546 (0.708)	35.612 (0.711)
	7% or 2,705 trimmed, N=35,936 **Uncorrected Gini: 45.490 (0.195)** **Gini using CPS wghts: 45.241 (0.233)**			**7% or 814 trimmed, N=10,820** **Uncorrected Gini: 34.753 (0.715)** **Gini using CAPMAS weights: 34.346 (0.605)**				
Gini corrected for unit non-response	45.718 (0.192)	46.421 (0.250)	46.296 (0.240)	35.826 (0.786)	35.819 (0.783)	35.783 (0.767)	35.800 (0.772)	35.707 (0.750)
Gini corrected for unit non-response & sampling wghts	45.484 (0.228)	46.192 (0.301)	46.063 (0.289)	35.457 (0.713)	35.390 (0.653)	35.531 (0.695)	35.542 (0.699)	35.588 (0.698)
	10% or 3,864 trimmed, N=34,777 **Uncorrected Gini: 45.477 (0.198)** **Gini using CPS wghts: 45.221 (0.234)**			**10% or 1,163 trimmed, N=10,471** **Uncorrected Gini: 34.907 (0.793)** **Gini using CAPMAS weights: 34.445 (0.650)**				
Gini corrected for unit non-response	45.722 (0.195)	46.008 (0.234)	45.979 (0.231)	35.933 (0.850)	35.926 (0.848)	35.943 (0.850)	35.957 (0.854)	35.848 (0.826)
Gini corrected for unit non-response & sampling wghts	45.475 (0.231)	45.742 (0.281)	45.713 (0.277)	35.507 (0.744)	35.445 (0.689)	35.632 (0.741)	35.639 (0.745)	35.673 (0.742)
	13% or 5,023 trimmed, N=33,618 **Uncorrected Gini: 45.479 (0.202)** **Gini using CPS wghts: 45.202 (0.240)**			**13% or 1,512 trimmed, N=10,122** **Uncorrected Gini: 34.795 (0.776)** **Gini using CAPMAS weights: 34.365 (0.651)**				
Gini corrected for unit non-response	45.786 (0.200)	45.889 (0.237)	45.781 (0.229)	35.842 (0.847)	35.798 (0.829)	35.801 (0.826)	35.828 (0.833)	35.699 (0.798)
Gini corrected for unit non-response & sampling wghts	45.532 (0.236)	45.609 (0.283)	45.501 (0.274)	35.470 (0.778)	35.350 (0.689)	35.543 (0.754)	35.564 (0.760)	35.585 (0.748)
	16% or 6,183 trimmed, N=32,458 **Uncorrected Gini: 45.537 (0.206)** **Gini using CPS wghts: 45.299 (0.246)**			**16% or 1,861 trimmed, N=9,773** **Uncorrected Gini: 34.713 (0.705)** **Gini using CAPMAS weights: 34.328 (0.609)**				
Gini corrected for unit non-response	45.853 (0.202)	45.682 (0.228)	45.642 (0.226)	35.668 (0.732)	35.697 (0.736)	35.707 (0.741)	35.732 (0.746)	35.622 (0.725)
Gini corrected for unit non-response & sampling wghts	45.631 (0.241)	45.432 (0.275)	45.394 (0.272)	35.317 (0.668)	35.305 (0.634)	35.478 (0.682)	35.496 (0.686)	35.532 (0.690)

Notes: Trimming of observations is randomized subject to household weights given by probability of response (equation 1) where $g=13-log(income)$. For clarity, Ginis and their standard errors are multiplied by 100. Ginis in columns 1-3 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 11. Standard errors on Ginis, in parentheses, are bootstrapped, and computed as per equation 12. The 22 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, 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 8. Semi-parametric estimates of Gini indexes: Pareto distribution for top 0.1–1% of incomes

Correction of extreme observ.	Sampling correction	2011 EU-SILC Observ. replaced	2011 EU-SILC Pareto coef. α	2011 EU-SILC Gini	2013 CPS Observ. replaced	2013 CPS α	2013 CPS Gini	2009 HIECS, 100% Observ. replaced	2009 HIECS, 100% α	2009 HIECS, 100% Gini
no (non-parametric estimation)	no			44.10			46.03			35.82
				(0.09)			(0.18)			(0.35)
	stat. agency weights			38.23			46.16			35.56
				(0.14)			(0.24)			(0.32)
	unit non-resp.			44.31			49.63			41.16
				(0.23)			(0.44)			(2.04)
	stat. agency weights & unit non-resp.			38.70			50.02			40.35
				(0.26)			(0.59)			(1.73)
yes (semi-parametric estimation) k=0.1%	no	214	2.087	44.10	73	3.288	46.03	46	2.033	35.82
			(0.144)	(0.14)		(0.377)	(0.19)		(0.361)	(0.30)
	stat. agency weights	193	1.989	38.23	59	3.706	46.16	49	2.066	35.56
			(0.186)	(0.19)		(0.620)	(0.24)		(0.340)	(0.37)
	unit non-resp.	91	1.654	44.31	16	5.407	49.63	9	0.810	41.17
			(0.193)	(0.34)		(0.755)	(0.44)		(0.141)	(8.47)
	stat. agency weights & unit non-resp.	71	2.041	38.70	11	22.740	50.02	12	0.901	40.36
			(0.278)	(0.34)		(12.970)	(0.59)		(0.183)	(5.11)
yes (semi-parametric estimation) k=0.2%	no	429	2.435	44.10	147	2.171	46.03	93	2.289	35.82
			(0.132)	(0.09)		(0.138)	(0.27)		(0.286)	(0.28)
	stat. agency weights	394	2.301	38.23	126	2.296	46.16	95	2.343	35.56
			(0.187)	(0.17)		(0.191)	(0.30)		(0.278)	(0.27)
	unit non-resp.	215	1.698	44.31	40	2.419	49.63	34	1.031	41.17
			(0.143)	(0.42)		(0.276)	(0.55)		(0.241)	(12.51)
	stat. agency weights & unit non-resp.	193	1.698	38.70	29	2.287	50.02	39	1.152	40.36
			(0.162)	(0.46)		(0.275)	(1.27)		(0.270)	(3.29)
yes (semi-parametric estimation) k=0.5%	no	1,072	2.875	44.10	368	2.325	46.03	234	2.720	35.82
			(0.104)	(0.08)		(0.116)	(0.23)		(0.216)	(0.25)
	stat. agency weights	993	2.728	38.23	333	2.178	46.16	240	2.723	35.56
			(0.153)	(0.14)		(0.135)	(0.46)		(0.204)	(0.28)
	unit non-resp.	632	2.137	44.31	134	1.890	49.63	132	1.469	41.16
			(0.128)	(0.28)		(0.139)	(0.71)		(0.308)	(1.32)
	stat. agency weights & unit non-resp.	576	2.096	38.70	103	2.020	50.03	140	1.588	40.35
			(0.165)	(0.28)		(0.195)	(0.72)		(0.307)	(0.98)
yes (semi-parametric estimation) k=1.0%	no	2,145	3.116	44.10	737	2.272	46.03	468	2.471	35.82
			(0.078)	(0.08)		(0.080)	(0.22)		(0.118)	(0.27)
	stat. agency weights	2,224	2.839	38.23	659	2.290	46.16	469	2.512	35.56
			(0.108)	(0.13)		(0.106)	(0.27)		(0.119)	(0.27)
	unit non-resp.	1,386	2.455	44.31	346	1.775	49.64	315	1.749	41.15
			(0.105)	(0.13)		(0.103)	(0.65)		(0.267)	(0.76)
	stat. agency weights & unit non-resp.	1,321	2.364	38.70	295	1.701	50.04	327	1.841	40.34
			(0.137)	(0.46)		(0.124)	(0.87)		(0.251)	(0.77)
Sample size (households)			214,581			73,765			46,857	

Notes: Pareto coefficients are estimated using maximum-likelihood methods. Semi-parametric Gini coefficients are computed as in equations 6 and 7. Their standard errors, in parentheses, are jackknife estimates and are computed using 30 random draws from the estimated Pareto distribution as in equation 13. Unit non-response bias is corrected using geographic disaggregation at the level of EU member states, US states, and Egyptian governorate urban–rural areas. EU-SILC sample is for 27 member states, excluding Croatia, Ireland, Portugal and Switzerland. For clarity, Ginis and their standard errors are multiplied by 100.

Table 9. Parametric & semiparametric estimates of Ginis: Generalized beta distribution

	EU-SILC (2011)				US CPS (2013)				HIECS (2009), 100% sample			
	No sampling correction	Stat. agency weights	Unit non-resp.	Stat. weight & unit non-resp.	No sampling correction	Stat. agency weights	Unit non-response	Stat. weight & unit non-resp.	No sampling correction	Stat. agency weights	Unit non-resp.	Stat. weight & unit non-resp.
E(a)	1.051	4.372	1.501	4.947	2.112	2.107	2.325	2.337	3.054	3.164	3.424	3.529
	(0.029)	(0.121)	(0.049)	(0.150)	(0.046)	(0.054)	(0.060)	(0.073)	(0.100)	(0.103)	(0.141)	(0.136)
E(b)	78,435.70	23,417.00	37,983.05	23,080.67	33,746.99	35,785.83	31,065.35	32,689.51	2,563.605	2,582.503	2,610.804	2,626.116
	(7,872.67)	(200.84)	(1,938.18)	(194.67)	(469.81)	(625.42)	(472.98)	(639.78)	(47.020)	(44.728)	(42.362)	(39.980)
E(p)	1.541	0.302	0.965	0.278	0.695	0.688	0.629	0.618	1.945	1.844	1.634	1.561
	(0.060)	(0.009)	(0.042)	(0.009)	(0.021)	(0.024)	(0.021)	(0.025)	(0.122)	(0.114)	(0.115)	(0.103)
E(q)	8.814	0.776	3.309	0.610	1.245	1.257	0.921	0.910	0.755	0.730	0.610	0.596
	(0.852)	(0.031)	(0.266)	(0.027)	(0.045)	(0.054)	(0.040)	(0.049)	(0.032)	(0.031)	(0.035)	(0.031)
Log(pseudo-likel.)	-2,288,898	-2.186×10^{9}	-2,972,056	-2.878×10^{9}	-832,897.1	-1.368×10^{9}	-929,569.5	-1.534×10^{9}	-429,258.4	-1.588×10^{8}	-449,834.3	-1.662×10^{8}
Parametric Gini	44.02	37.79	43.59	37.98	46.03	46.10	49.36	49.64	35.85	35.58	38.35	37.96
	(0.07)	(0.11)	(0.10)	(0.16)	(0.17)	(0.22)	(0.36)	(0.47)	(0.23)	(0.23)	(0.47)	(0.42)
Semiparam. Gini, top	43.69	37.95	43.46	38.13	45.94	46.02	49.65	50.05	35.79	35.50	38.33	38.09
0.1% replaced	(0.06)	(0.12)	(0.09)	(0.18)	(0.20)	(0.24)	(0.54)	(0.68)	(0.31)	(0.31)	(0.88)	(0.78)
Semiparam. Gini, top	43.63	37.92	43.33	38.01	45.87	45.90	49.48	49.89	35.80	35.62	38.23	38.00
0.2% replaced	(0.06)	(0.12)	(0.08)	(0.17)	(0.21)	(0.23)	(0.57)	(0.71)	(0.30)	(0.30)	(0.64)	(0.72)
Semiparam. Gini, top	43.60	37.88	43.25	37.96	45.82	45.85	49.04	49.42	35.86	35.72	38.43	38.05
0.5% replaced	(0.06)	(0.12)	(0.08)	(0.17)	(0.19)	(0.22)	(0.42)	(0.61)	(0.29)	(0.36)	(0.86)	(0.62)
Semiparam. Gini, top	43.63	37.88	43.26	37.93	45.80	45.89	49.11	49.22	35.87	35.57	38.28	37.85
1% replaced	(0.06)	(0.12)	(0.08)	(0.16)	(0.19)	(0.26)	(0.58)	(0.60)	(0.30)	(0.31)	(0.57)	(0.50)
Semiparam. Gini, top	43.74	37.90	43.35	37.95	45.89	45.94	49.00	50.05	35.86	35.74	38.34	37.88
2% replaced	(0.06)	(0.11)	(0.08)	(0.15)	(0.20)	(0.23)	(0.41)	(0.68)	(0.33)	(0.43)	(0.75)	(0.49)
Semiparam. Gini, top	44.00	37.93	43.60	38.03	45.98	46.07	49.40	49.39	35.92	35.55	38.09	37.82
5% replaced	(0.06)	(0.11)	(0.07)	(0.14)	(0.19)	(0.25)	(0.34)	(0.45)	(0.33)	(0.29)	(0.40)	(0.42)
Semiparam. Gini, top	44.21	37.97	43.79	38.10	46.00	46.12	49.29	49.53	35.91	35.62	38.29	37.76
10% replaced	(0.06)	(0.11)	(0.07)	(0.15)	(0.19)	(0.22)	(0.33)	(0.44)	(0.34)	(0.30)	(0.40)	(0.38)
Semiparam. Gini, top	44.28	37.98	43.87	38.10	45.94	45.97	49.31	49.56	35.78	35.61	38.32	37.98
20% replaced	(0.06)	(0.11)	(0.07)	(0.14)	(0.19)	(0.22)	(0.39)	(0.39)	(0.28)	(0.32)	(0.52)	(0.41)

Notes: Standard errors are in parentheses. Parametric Ginis are calculated by numerical integration with 5,000 integration points. Semi-parametric Ginis are computed as in equations 7 and 12. Standard errors of semiparametric Ginis, in parentheses, are jackknife estimates and are computed using 30 random draws from the estimated generalized beta type-2 distribution as in equation 13. EU-SILC sample is for 27 member states, excluding Croatia, Ireland, Portugal and Switzerland. For clarity, Ginis and their standard errors are multiplied by 100.

5. Conclusions

This study has evaluated several methods for correcting of statistical problems with top incomes, including unit non-response and representativeness of top income observations. The joint use of two distinct statistical methods for correcting top incomes biases, sensitivity analysis of their technical specifications, and analysis of their performance on three vastly different household surveys were methodological contributions of this study. The European Union Statistics on Income and Living Conditions, the United States Current Population Survey and the Egyptian Household Income, Expenditure and Consumption Survey were used as prototypes of worldwide surveys with different types of measurement issues. We first tested for the problem of unit non-response 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 in this paper suggests that unit non-response is responsible for a significant 0.4–9.7 percentage point bias in the Gini index of inequality in the US CPS, a 0.9–5.3 percentage point bias in the Egyptian HIECS, but only a modest 0.1–0.5 percentage point bias in the EU-SILC. This divergence stems from several differences between the three respective datasets. In the case of the HIECS data, the non-response bias correction is limited by the low observed non-response rate and by homogeneity of households within PSUs, which prevent the model from estimating response probabilities too low. In other national surveys, such as the US CPS, response probabilities can be estimated very low for some households, because other households in the same region, of different demographics, can be assigned very high probabilities in compensation.

In the EU-SILC, the low correction may also be attributed to relatively little overlap in the income distributions of various member states. The narrow range of estimates for the EU-SILC, rather than implying precision of estimation, reflects on limitations in the ways EU-SILC data can be analyzed. Income distributions vary significantly across member states with relatively little overlap. Economic and cultural differences across member states also put the assumption of stability of behavioral responses across regions into question, suggesting that we may not estimate a clear response-probability function. Data on unit non-response rates at lower levels of geographic aggregation – at which the assumption of behavioral stability is more likely to hold – are missing.

The second most significant finding of this study is that changing of the geographic level of analysis has an important systematic impact on the unit non-response correction. Greater degrees of geographic disaggregation typically yield lower estimates of the non-response bias, but the bias remains significant. The degree of geographic disaggregation is thus an important parameter to consider in correcting for unit non-response through reweighting. That implies that understanding of the income distribution, demographics and behavioral similarities in the population within and across regions is important. An experiment on two high quality samples suggested that a medium degree of disaggregation achieves the best estimate of the bias and correction for it.

Correcting for non-representative distributions of top income observations using fitted values or random draws from the Pareto or generalized beta distributions helps to refine the estimated Gini, but by a small magnitude. In the EU-SILC and the Egyptian HIECS the correction was downward, of up to 0.014 percentage points, and suggested that the observed top 0.1% of incomes may be extreme or overstated, commanding an undue share of national income, while the following 1% of incomes followed typical distributions more closely. In the US CPS, on the other hand, the correction was either negative or positive, depending on whether generalized beta distribution or Pareto distribution was applied, respectively. Using the Pareto approximation, income share of the super-rich 0.1% of households is estimated to be not as high as in other income distributions or under a smooth Pareto curve, but the income share of the next 1% of incomes is higher. That may serve as a confirmation that topmost incomes in the US CPS are top-coded, or may suggest that extreme observations appear among the top 1% of incomes, rather than among the super-rich 0.1%. In any case, the assumption regarding the true distribution of top incomes has a small effect on the correction, particularly relative to the correction for unit non-response.

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.

Bates, N. and Creighton, K. (2000) The last five percent: What can we learn from difficult interviews?, Proceedings of the Annual Meetings of the American Statistical Association, August 13-17, 2000. http://www.fcsm.gov/committees/ihsng/asa2000proceedings.pdf.

Belhaj Hassine, N., (2011). Inequality of Opportunity in Egypt. World Bank Economic Review, 26(2), 265 – 295.

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

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 Jenkins, S.P. (2009) Using the P90/P10 ratio to measure US inequality trends with Current Population Survey data: a view from inside the Census Bureau vaults, Review of Income and Wealth, 55, 166-185.

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.

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., and Verme, P. (2013) Top Incomes and the Measurement of Inequality in Egypt, World Bank Policy Research working paper series #6557.

Hokayem, C., Ziliak, J.P., and Bollinger, C.R. (2012) A look at CPS non-response and trends in poverty, US Census SEHSD Working Paper 2012-21.

Jenkins, S.P. (2009) GB2LFIT: Stata module to fit a GB2 distribution to unit record data, Institute for Social and Economic Research, University of Essex, Colchester, UK.

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.

Jolliffe, D., Datt, G., and Sharma, M. (2004) Robust poverty and inequality measurement in Egypt: Correcting for spatial-price variation and sample design effects, Review of Development Economics, 8(4), 557-572.

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.

Killewald, A., Andreski, P., and Schoeni, R. (2011) Trends in Item Nonresponse in the PSID 1968-2009.

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.

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.

Neri, L., Gagliardi, F., Ciampalini, G., Verma, V. and Betti, G. (2009) Outliers at upper end of income distribution (EU-SILC 2007), DMQ Working Paper n. 86, November 2009.

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.

Platek, R. (1977) Some factors affecting non-response. Survey Methodology. 3:191-214. Bulletin of the International Statistical Institute.- [Wechselnde Verlagsorte], ISSN 0373-0441, ZDB-ID 2150426. 47.1977, 3, 347-366.

Proctor, B.D. and Dalaker, J. (2002) Poverty in the United States: 2001, U.S. Census Bureau, Current Population Report P60-219, U.S. Government Printing Office, Washington, DC.

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.

Tiehen, L. Jolliffe, D. and Gundersen, C. (2013) Poverty and food assistance during the great recession, working paper.

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

World Bank (2013) Inside Inequality in Egypt: Historical trends, recent facts, people's perceptions and the spatial dimension, mimeo.

Annex

Table A1. Non-response rate and income distribution by member state, 2009 EU-SILC

Member State	Households	Non-response Rate (%)	Mean Equivalised Disposable Income per Capita (Euro)	Member State Gini, EU-SILC weighted households
Austria	5,875	28.1	22,186.58	26.99
Belgium	6,107	36.7	21,114.74	27.10
Bulgaria	5,583	22.5	3,245.85	34.82
Cyprus	3,144	10.5	19,130.27	32.19
Czech Republic	9,908	17.7	8,210.21	26.02
Denmark	5,811	46.5	26,279.67	24.92
Estonia	4,952	25.4	7,113.63	33.10
Finland	10,128	20.8	22,845.36	27.65
France	10,597	16.9	23,382.75	30.40
Germany	13,065	23.1	21,112.46	30.42
Greece	6,951	15.2	13,606.02	32.69
Hungary	9,907	15.4	5,237.55	24.63
Iceland	2,893	26.9	26,452.07	30.75
Ireland	5,174	21.1	25,678.21	30.05
Italy	20,363	16.3	18,156.96	31.68
Latvia	5,760	20.8	6,369.70	38.79
Lithuania	5,103	13.0	5,815.69	36.12
Luxembourg	4,243	48.1	36,985.05	29.32
Malta	3,645	20.2	11,941.52	28.20
Netherlands	9,708	16.6	22,883.81	27.13
Norway	5,423	39.6	35,940.48	25.68
Poland	13,221	17.4	6,019.32	32.25
Portugal	4,961	13.1	10,407.29	36.01
Romania	7,670	3.5	2,552.65	34.44
Slovakia	5,256	11.5	6,277.28	25.08
Slovenia	9,281	22.3	12,597.30	24.75
Spain	13,153	17.9	14,880.70	31.92
Sweden	7,510	27.0	22,485.91	26.02
Switzerland	7,357	24.8	34,443.89	31.08
United Kingdom	8,314	28.7	19,496.29	32.54
Wtd. Mean (Total)	7,702 (231,063)	22.24	17,485.22	30.73 (38.16)

Note: Non-response rate is reported in the member-states' Intermediate/Final Quality Reports at the state level as *NRh* for total sample. All states from table 1 plus Ireland, Portugal and Switzerland are included. (Croatia was omitted from the EU-SILC survey until the 2010 wave.) Per-capita income is weighted by household size. Incomes less than 1 are omitted. Mean incomes may not be representative of those for the entire states, as they omit non-responding households. For clarity of presentation, Ginis are multiplied by 100.

Table A2. Non-response rate and income distribution by governorate, 2009 HIECS (25%)

Governorate	PSUs	Households	Non-response Rate (%)	Mean Income per Capita (E)	Governorate Gini, CAPMAS-Weighted Hhds.
Alexandria	148	694	6.0	5,347.73	32.88
Assiut	100	459	2.4	2,746.14	35.40
Aswan	52	236	1.0	3,597.35	28.59
Behera	152	704	0.6	3,620.25	23.83
Beni Suef	67	295	1.3	2,835.28	24.92
Cairo	284	1,308	8.9	6,651.74	40.51
Dakahlia	176	756	1.6	4,456.14	27.79
Damietta	52	248	2.9	5,567.66	28.78
Fayoum	78	394	1.1	3,041.00	22.80
Gharbia	139	653	2.2	4,461.02	27.75
Giza	214	993	6.5	4,537.38	39.56
Ismailia	52	252	2.1	6,260.87	50.21
Kafr ElSheikh	85	405	4.2	4,424.67	27.05
Kalyoubia	144	658	3.2	4,252.08	28.82
Luxor	14	72	1.1	5,332.94	35.26
Matrouh	11	56	0.0	5,195.80	24.36
Menia	128	591	2.5	3,561.99	34.11
Menoufia	106	477	2.8	3,988.81	31.13
New Valley	8	39	3.9	5,220.90	29.73
North Sinai	14	70	10.5	3,683.03	25.59
Port Said	49	204	7.4	6,333.59	34.16
Qena	87	415	2.6	3,432.64	30.74
Red Sea	13	59	3.2	6,646.68	29.89
Shrkia	175	793	1.9	3,610.43	24.55
South Sinai	4	25	9.2	12,662.86	78.45
Suez	50	242	4.9	7,490.56	37.08
Suhag	113	536	1.0	2,837.55	27.20
Wtd. Mean (Total)	93 (2,515)	431 (11,634)	3.7	4,321.06	31.51 (36.01)

Notes: Non-response rate, reported in the survey at the PSU level, is weighted by the number of responding households in each PSU. Per-capita income and expenditure are further weighted by household size. Mean incomes may not be representative of those for the entire governorates, as they omit non-responding households. For clarity, Ginis are multiplied by 100.

Table A3. Semi-parametric estimates of Gini indexes: Pareto distribution for top 2–20% of incomes

Correction of extreme observ.	Sampling correction	2011 EU-SILC Observ. replaced	Pareto coef. α	Gini	2013 CPS Observ. replaced	α	Gini	2009 HIECS, 100% Observ. replaced	α	Gini
	no			44.10			46.03			35.82
				(0.09)			(0.18)			(0.35)
no (non-	stat. agency			38.23			46.16			35.56
parametric	weights			(0.14)			(0.24)			(0.32)
estimation)	unit non-resp.			44.31			49.63			41.16
				(0.23)			(0.44)			(2.04)
	stat. agency			38.70			50.02			40.35
	weights &			(0.26)			(0.59)			(1.73)
	unit non-resp.									
	no	4,291	3.324	44.10	1,475	2.378	46.03	937	2.369	35.82
			(0.058)	(0.07)		(0.063)	(0.23)		(0.076)	(0.29)
yes (semi-	stat. agency	4,766	3.018	38.22	1,358	2.301	46.16	969	2.346	35.56
parametric	weights		(0.083)	(0.12)		(0.076)	(0.29)		(0.074)	(0.32)
estimation)	unit non-resp.	2,987	2.783	44.30	848	1.769	49.65	713	1.822	41.13
			(0.082)	(0.10)		(0.075)	(0.97)		(0.172)	(0.69)
k=2%	stat. agency	3,241	2.547	38.69	740	1.703	50.05	737	1.860	40.32
	weights &		(0.101)	(0.26)		(0.093)	(1.11)		(0.153)	(0.60)
	unit non-resp.									
	no	10,729	3.377	44.09	3,688	2.436	46.03	2,342	2.350	35.82
			(0.035)	(0.07)		(0.041)	(0.20)		(0.049)	(0.31)
yes (semi-	stat. agency	11,973	3.231	38.21	3,329	2.476	46.15	2,438	2.326	35.56
parametric	weights		(0.057)	(0.12)		(0.054)	(0.25)		(0.048)	(0.31)
estimation)	unit non-resp.	8,099	3.017	44.27	2,595	1.959	49.63	1,984	1.913	41.05
			(0.049)	(0.09)		(0.055)	(0.45)		(0.094)	(0.86)
k=5%	stat. agency	9,303	2.827	38.67	2,306	1.885	50.04	2,092	1.948	40.26
	weights &		(0.069)	(0.15)		(0.069)	(0.93)		(0.084)	(0.51)
	unit non-resp.									
	no	21,458	3.159	44.10	7,376	2.409	46.04	4,679	2.307	35.83
			(0.021)	(0.08)		(0.028)	(0.21)		(0.033)	(0.33)
yes (semi-	stat. agency	22,487	3.228	38.20	6,610	2.442	46.15	4,876	2.282	35.57
parametric	weights		(0.039)	(0.12)		(0.036)	(0.27)		(0.033)	(0.32)
estimation)	unit non-resp.	17,102	2.956	44.24	5,797	2.032	49.59	4,235	2.033	40.87
			(0.030)	(0.09)		(0.039)	(0.44)		(0.062)	(0.47)
k=10%	stat. agency	18,823	2.910	38.62	5,147	2.014	49.96	4,378	2.032	40.12
	weights &		(0.047)	(0.14)		(0.051)	(0.44)		(0.055)	(0.50)
	unit non-resp.									
	no	42,916	2.736	44.35	14,753	2.215	46.27	9,367	2.213	35.93
			(0.012)	(0.09)		(0.017)	(0.35)		(0.022)	(0.37)
yes (semi-	stat. agency	41,058	2.990	38.28	13,467	2.208	46.41	9,695	2.212	35.67
parametric	weights		(0.023)	(0.14)		(0.021)	(0.30)		(0.022)	(0.32)
estimation)	unit non-resp.	36,061	2.659	44.37	12,637	1.969	49.69	8,829	2.033	40.60
			(0.016)	(0.11)		(0.024)	(0.38)		(0.037)	(0.44)
k=20%	stat. agency	35,915	2.838	38.61	11,471	1.908	50.18	9,141	2.043	39.89
	weights &		(0.029)	(0.15)		(0.029)	(0.55)		(0.034)	(0.49)
	unit non-resp.									
Sample size (households)			214,581			73,765			46,857	

Notes: Pareto coefficients are estimated using maximum-likelihood methods. Semi-parametric Gini coefficients are computed as in equations 6 and 7. Their standard errors are computed using 30 random draws from the estimated Pareto distribution as in equation 13. Jackknife standard errors are in parentheses. Unit non-response bias is corrected using geographic disaggregation at the level of EU member states, US states, and Egyptian governorate urban–rural areas. EU-SILC sample is for 27 member states, excluding Croatia, Ireland, Portugal and Switzerland. For clarity, Ginis and their standard errors are multiplied by 100.