

Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods Using the EU SILC Data

Hlasny, Vladimir; Verme, Paolo

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Hlasny, V., & Verme, P. (2018). Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods Using the EU SILC Data. *Econometrics*, 6(2), 1-21. <https://doi.org/10.3390/econometrics6020030>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Article

Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods Using the EU SILC Data

Vladimir Hlasny ^{1,*}  and Paolo Verme ²¹ Department of Economics, Ewha Womans University, Seoul 03760, Korea² The World Bank, Washington, DC 20433, USA; pverme@worldbank.org

* Correspondence: vhlasny@ewha.ac.kr

Received: 1 January 2018; Accepted: 23 May 2018; Published: 4 June 2018



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 top-income biases generated by data issues such as unit or item non-response. Results for the European Union's Statistics on Income and Living Conditions survey indicate that survey response probabilities are negatively associated with income and bias the measurement of inequality downward. Correcting for this bias with *reweighting*, the Gini coefficient for Europe is revised upwards by 3.7 percentage points. Similar results are reached with *replacing* of top incomes using values from the Pareto distribution when the cut point for the analysis is below the 95th percentile. For higher cut points, results with *replacing* are inconsistent suggesting that popular parametric distributions do not mimic real data well at the very top of the income distribution.

Keywords: top incomes; inequality measures; survey non-response; Pareto distribution; parametric estimation; EU SILC

JEL Classification: D31; D63; N35

1. Introduction

Thanks to the wide public attention that top incomes have received in the aftermath of the global financial crisis, it is now acknowledged that top incomes have grown disproportionately faster than other incomes in industrialized countries over the past several decades. The fact that these top incomes are difficult to capture in household surveys potentially leads to biases in the estimation of income inequality related to the representation and precision of reported top incomes, even though the direction of the bias is not a priori clear (Deaton 2005, p. 11). These range from issues related to sampling, to issues related to data collection, data preparation or data analysis. The European Union Survey of Income and Living Conditions, for example, suffers from data issues such as under-representation of the highest incomes (Bartels and Metzing 2017; Törmälehto 2017). Most countries in Europe suffer from very high non-response rates reaching up to 50 percent of the sample. Income measurement issues including surveying, interview methods and post-survey treatment also explain differences in inequality measurements across data sources (Frick and Krell 2010).

Two types of in-survey methods have been proposed to address the question of correcting inequality in the presence of top-income biases while relying on survey microdata only. The first method, which we call *reweighting*, attempts to correct the sampling weights of existing observations using information on unit or item non-response rates across demographic cells such as geographical

areas (Mistiaen and Ravallion 2003; Korinek et al. 2006, 2007). The approach exploits the relationship between response rates and shapes of income distributions across national regions to estimate the gradient of households' response probability by income level. It then uses the estimated response probabilities to reweight the observed incomes by the mass of nonresponding households in order to correct the measure of inequality. The second method, which we call *replacing* attempts to replace top-income observations with observations generated from known theoretical distributions. This method can be used to correct for issues such as top coding, trimming or censoring but can also mitigate the problem of unit or item non-responses if these non-responses are concentrated among top incomes (Cowell and Victoria-Feser 2007; Jenkins et al. 2011). Several distributions have been suggested as candidates, including Pareto type I or type II, or generalized beta.¹ Hlasny and Verme (2018) have combined the reweighting and replacing methods, and studied the contribution of each method to the composite correction of an inequality index.

It is evident that both the reweighting and replacing methods have their advantages and disadvantages, as the information available within surveys has its limits even if used creatively to correct for top-income problems. Proper reweighting and replacing depend on the appropriateness of parametric assumptions imposed on a particular national distribution of incomes at hand. Using alternative methods based on out of survey information such as tax records or national accounts data to inform the measurement of top incomes has its own measurement problems. Good tax or macro data are only available in a few countries and data may not be comparable across countries, whereas household survey data of reasonable quality are now available in most countries worldwide.

This paper compares the reweighting and replacing methods using the European Union's Statistics on Income and Living Conditions (SILC) survey data, taking into account heterogeneity of income distributions, differences in sampling designs and definitions of non-response rates across EU member states. We find survey non-response probabilities to be negatively and significantly associated with income indicating that measures of inequality are downward biased. Correcting for this bias with reweighting, the Gini coefficient for Europe is revised upwards by 3.7 percentage points. Similar results are reached with replacing of top incomes using values from the Pareto distribution when the cut point for replacing is set below the 95th percentile. For higher cut points, results with replacing are inconsistent suggesting that popular parametric distributions do not mimic well real data at the very top of the income distribution.

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 3 describes the data. Section 4 presents main results and Section 5 concludes.

2. Materials and Methods

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 clarifying the type of error we address in this paper.

Sample design issues 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. It is important to

¹ Similar methods include Lakner and Milanovic (2013) who combined corrections for unit non-response with corrections for measurement errors among top incomes, and calibrated the estimated Pareto distribution among top incomes using aggregate income information from national accounts data. Bartels and Metzger (2017) replaced the top one percent of incomes in the EU Statistics on Income and Living Conditions (SILC) surveys with Pareto estimates obtained using World Wealth and Income Database information.

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 3000 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 of euros 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.² 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 using rereighting.

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-response biases results if it is non-random and related to specific factors. Non-response may be related to households' characteristics such as wealth or education, and this may bias statistics constructed with income or expenditure variables. 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) or by means of imputation using household and individual socio-economic characteristics to predict income. The rereighting method proposed in this paper also corrects for item 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. The replacing method used in this paper helps to correct for item underreporting.

Generic measurement errors. Any variable including income or expenditure can be subject to measurement error. This error is typically expected to be random, distributed 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 explicitly but these errors are implicitly treated when replacing observations.

Data preparation issues 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.

² Notable exception is that of income surveys based on national tax registers (Burricand 2013; Jäntti et al. 2013).

Topcoding. Top coding is the practice adopted by some statistical agencies such as the US Census Bureau to modify intentionally the values of some variables to prevent identification of households or individuals. It can take various forms, from replacing values above a certain threshold with means or medians of top cells to swapping incomes across top observations. In some cases and for research purposes, statistical agencies provide restricted access to the original values. However, in most cases researchers are left with the problem of having to correct sample statistics for top coding. In this paper, we use EU-SILC data which are not subject to top coding on the part of Eurostat, although it is possible that some countries apply some form of topcoding to their data before transmitting these data to Eurostat. Replacing corrects for topcoding but only for the segment of data replaced whereas reweighting is unlikely to correct for topcoding.

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 or to limit their influence if their variable values may have been mismeasured. The overarching objective is to control the influence of units that are rare in the sampling frame. Trimming observations or weights biases statistical measurement and should be corrected for. Trimming is similar to unit or item non-response in that we are missing income observations. Reweighting can help to address this issue if trimmed income observations come from within the support in the observed sample.

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 (2018) have tested the margins of error in inequality measurement that can arise from the provision of subsamples instead of full samples and found significant margins of error. This issue is not treated in this paper because EU-SILC data are provided in full.

Data analysis issues 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). Based on these findings, 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 of the implications for the measurement of inequality.

2.1. Reweighting

Unlike the case of item non-response, unit non-response cannot be dealt with by inferring households’ unreported income from their other reported characteristics, because we do not observe any information for the non-responding households. In an effort to address this problem, Atkinson and Micklewright (1983) used information on non-response rates across regions to uniformly ‘gross up’ the mass of respondents in a region by the regional non-response rate. This is the approach taken by several national statistical agencies in adjusting sampling weights for regional unit non-response. This approach is inadequate, as it accounts only for inter-regional differences in non-response rates, and not for systematic differences in response probability across units within individual regions.

Mistiaen and Ravallion (2003), and Korinek et al. (2006, 2007) proposed a probabilistic model that uses information on non-response rates across geographic regions as well as information about the distribution within regions. They estimated the response probability of each household, and used the inverse of this estimate to adjust each household's weight. Each household's weight is thus 'grossed up' non-uniformly to match the mass of all respondents to the size of the underlying population.

The central tenet of the method is that the probability of a household i in a region j to respond to the survey, P_{ij} , is a deterministic function of its arguments. Logistic functional form is used for its simplicity and its robustness properties:

$$P_{ij}(x_{ij}, \theta) = \frac{e^{g(x_{ij}, \theta)}}{1 + e^{g(x_{ij}, \theta)}} \quad (1)$$

Here $g(x_{ij}, \theta)$ is a stable function of x_{ij} , the observable demographic characteristics of responding households that are used in estimations, and of θ , the corresponding vector of parameters. Variable-specific subscripts are omitted for conciseness. $g(x_{ij}, \theta)$ is assumed to be twice continuously differentiable. Equation (1) thus imposes several restrictions on the modeled behavioral relationship between households' characteristics x_{ij} and their response probability: the relationship is deterministic and dictated by the logistic functional form and the functional form of $g(x_{ij}, \theta)$, differentiable at all levels of x_{ij} , and identical across all households and regions. These restrictions are strong, but several facts help to justify them. One, the logistic function is well-accepted as a robust form to model probabilistic relations. Two, Korinek et al. (2006, 2007), and Hlasny and Verme (2018) have evaluated alternative forms of $g(x_{ij}, \theta)$ including non-monotonic functions on US and Egyptian data, and have concluded that some of the most parsimonious functions provide very good fit, compared to both uncorrected income distributions and compared to external information on the true degree of inequality in those countries. Three, nonlinear forms of $P(x_{ij}, \theta)$ and $g(x_{ij}, \theta)$ allow for response differences between poorer and richer households in a realistic way. Four, a comparative study of US, EU and Egyptian data led to similar estimation results across countries, suggesting that the behavioral tendencies exhibit a high degree of consistency across regions (Hlasny and Verme 2015). Five, supplementing $g(x_{ij}, \theta)$ with indicators for subsets of regions helps to attenuate any systematic behavioral differences across parts of the country.

The number of households in each region (\hat{m}_j) is 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.

$$\hat{m}_j = \sum_{i=1}^{N_j} \hat{P}_{ij}^{-1}(x_{ij}, \hat{\theta}) \quad (2)$$

The parameters θ can be estimated by fitting the estimated and actual number of households in each region using the generalized method of moments estimator:

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_j \left[(\hat{m}_j - m_j) w_j^{-1} (\hat{m}_j - m_j) \right] \quad (3)$$

where m_j is the number of households in region j according to sample design, and w_j is a region-specific analytical weight proportional to m_j .³ The asymptotic variance of $\hat{\theta}$ can be estimated as the ratio of the

³ An illustration is in order. Suppose there are two income groups residing in two national regions. Region 1 has a higher share of the richer income group, and correspondingly a higher unit non-response rate, as the richer households are less likely to participate. As a result, mean income and income inequality index may or may not differ across the two regions. To correct the mean incomes and inequality indexes in each region as well as nationally, we wish to give more weight to each richer household until the sum of weights equals the underlying regional population, because behind each responding rich household there are more non-responding rich households. Equation (2) 'blows up' the weight of each responding household systematically, under the household-level behavioral rules specified in Equation (1), to fit the joint weighted mass of the responding households to the underlying regional population (Equation (3)). In one region the weighted mass of the responding households may exceed the underlying population, while in the other region it may fall short (because of the restrictions imposed in Equation (1)), but the nationwide sum of the weighted masses equals the underlying national population.

model objective value (the weighted sum of squared region-level residuals), and the squared partial derivative of this objective value with respect to $\hat{\theta}$ (equal to $-\sum_j w_j^{-1} \sum_i (x_{ij} / e^{x_{ij}\hat{\theta}})$ under the assumed logistic functional form), both weighted by region-specific analytical weights w_j (Equations (11)–(14) in Korinek et al. 2007).

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 $g(x_{ij}, \theta)$ 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.

Applying the model in Equations (1)–(3) involves making several decisions regarding the delineation of regions, and choosing parametric forms for the functions $P(x_{ij}, \theta)$ and $g(x_{ij}, \theta)$. The choice of regional delineation involves a trade-off between the number of j data points for the model loss function (Equation (3)), and the number and distribution of within- j observations vis-à-vis the underlying population to achieve consistency for the underlying distribution of incomes. The sample in each region should encompass the entire range of values of relevant characteristics of the underlying population, calling for a higher geographic level at which sample stratification was performed.

Properties of the data at hand thus call for different degrees of data aggregation, but there is presently little guidance for arbitrary national surveys. For the United States CPS, Korinek et al. (2006, 2007) used state-level aggregation, because geographic identifiers are consistently reported only at that level whereas county or metropolitan statistical area identifiers are missing for some responding as well as non-responding households. Hlasny and Verme (2017) considered various degrees of geographic aggregation, from the level of 185 metropolitan statistical areas (MSAs) to that of 7 census divisions. They concluded that an intermediate level of aggregation, at the level of states or groups of 1–2 MSAs, performed more consistently than extreme aggregation or disaggregation. Using the Egyptian Household Income, Expenditure and Consumption Survey (HIECS), Hlasny and Verme (2018) assessed the degrees of regional aggregation from a high administrative level (governorate by urban–rural areas, 50 areas with 939.7 observations on average) down to the level of primary sampling units (PSUs, 2526 areas with 18.6 observations on average). These alternative approaches yielded different corrections for unit non-response, but the more detailed level of disaggregation was deemed conceptually more appropriate. It gave rise to a higher number of data points used in optimization (Equation (3)). Moreover, the observed range of household characteristics in each Egyptian PSU likely comprised the values of non-responding households, while higher levels of geographic aggregation would make behavioral responses less stable across households within areas j .

For the set of national surveys in the SILC, this paper uses regional aggregation to the highest level of nomenclature of territorial units for statistics (NUTS-1) level. With the exception of a handful of countries, non-response rates are not available at more detailed levels of disaggregation. At the same time, heterogeneity of non-response rates reported by national statistical agencies puts aggregation to the level of EU member states into question. In a similar vein, to satisfy the assumption of stability of $g(x_{ij}, \theta)$ across all regions, functional form and covariates x_{ij} are selected to make households across all regions behaviorally similar, in the sense that households with similar values of demographic variables should have similar response probabilities across all regions. To effectively neutralize the cross-country heterogeneity in households' response probabilities, logarithmic specification of $g(x_{ij}, \theta)$, and country indicators are used in $g(x_{ij}, \theta)$. On the margins, we will report how the addition of regional indicators affects the correction for the unit non-response bias.⁴

⁴ Exclusion of influential regions and EU member states was also tried, but is not discussed here, as it prevents the estimation of inequality for EU member states and EU at large. (These results are available on request.)

For the covariates in x_{ij} , Korinek et al. (2006, 2007) evaluated several variables affecting households' response probability, including income, gender, race, age, education, employment status, household size and an urban–rural indicator. Hlasny and Verme (2018) compared income and expenditures, and indicators for survey rounds. These studies concluded that univariate models controlling for expenditures or income are the most efficient. Because this paper focuses on equivalized disposable income as the welfare aggregate, and because arbitrary household surveys worldwide may not consistently report any additional household characteristics, equivalized disposable income is used as the only explanatory variable.⁵

Finally worth noting, SILC surveys already provide a limited correction for unit non-response through sampling weights. This method accounts for differences in response rates across regions but not for systematic differences across demographic groups within regions. Unfortunately, these sampling weights cannot be decomposed into weights for unit non-response and weights for other issues with unit representativeness. We could either double-correct for unit non-response by using the available sampling weights, or ignore other sample representativeness issues by not using the weights. In the United States CPS (Korinek et al. 2006, 2007; Hlasny and Verme 2017) and the Egyptian HIECS (Hlasny and Verme 2018), the correction for non-response (through \hat{p}_i^{-1}) affected inequality estimates substantially more than the corrections for other sample representativeness issues (through sampling weights), and so the non-response correction weights should be used with or without the survey sampling weights. These findings may not apply to surveys with less prevalent or less systematic non-responses, and with graver sampling design issues. In the case of the SILC, the great heterogeneity in sample representativeness across EU member states, and the modest role of non-response correction in the available sampling weights are thought to favor the usage of the non-response correction weights (\hat{p}_i^{-1}) in tandem with the sampling weights. To accommodate all these options, alternative estimates of inequality are produced: on uncorrected data, data corrected with non-response-bias weights, data corrected with statistical agency weights, and data corrected with both sets of weights simultaneously. Estimates obtained without sampling weights are reported on the margins.

2.2. Replacing

An alternative approach to correct for poorly reported top incomes is to remove the top end of the distribution and replace it with synthetic values under some parametric assumptions. Cowell and Victoria-Feser (1996), Cowell and Flachaire (2007) and a large body of following studies combined estimates from a Pareto distribution (Pareto 1896) for the top of the income distribution with non-parametric statistics for the rest of the distribution. Atkinson et al. (2011) summarize this literature, and model the historic distribution of top incomes in several countries. Testing this method on US CPS data, Hlasny and Verme (2017) find that replacing actual top incomes with Pareto parametric estimates has a small positive effect on the computed Gini, implying that the reported top incomes are distributed more narrowly than the predicted values. However, the effect is smaller than a correction for unit non-response alone using the reweighting method, suggesting that top-income biases operating in opposite directions may be at play. Burkhauser et al. (2010) compared four alternative parametric estimators for replacing of topcoded incomes and combined the estimates with those from non-topcoded incomes. Alvaredo and Piketty (2014) have recently proposed to use synthetic data for the entire income distribution, and estimate inequality using a mix of Pareto distributions for top incomes and log-normal distributions for the rest of incomes. Alvaredo et al. (2017) improve on this methodology by collecting survey micro-data from several countries, and replacing top incomes with values from the Pareto distribution benchmarked using administrative income tax data from a highly unequal paragon country, Lebanon. Using uncharacteristically high parametric values for

⁵ This decision also can be viewed as upholding the anonymity axiom that inequality measures be based only on the welfare aggregate itself, and independent of other household characteristics (Litchfield 1999).

the distributions in the Middle East countries, these approaches yielded higher inequality measures than those using raw survey data or using the Pareto replacement of top incomes alone (estimated by Hlasny and Intini 2015; Hlasny and Verme 2018).

Beside the Pareto distribution, other parametric forms have been suggested in recent literature as providing superior fit to income distributions in particular countries. A generalized beta distribution of the second kind (GB2), also known as the Feller-Pareto distribution, is a suitable functional form representing well a large extent of the income distribution (McDonald 1984). The upper tail of the distribution can be modeled as heavy and decaying similar to a power function, while the lower end of the distribution can be short-tailed. The lognormal, Fisk, Singh-Maddala (Singh and Maddala 1976) and Dagum (1980) distributions have also been suggested as candidates for modeling income distributions, being themselves limiting cases of the GB2 distribution with some parameters held fixed (McDonald 1984). However, their fit was not consistent or universally good across various waves of European and US income surveys (Butler and McDonald 1989; Brachmann et al. 1996; Jenkins 2007; Jenkins 2009a; Brzezinski 2013), and so the more flexible GB2 distribution may be preferred.

This study uses the parametric properties of the Pareto and GB2 distributions to evaluate how representative are the top-income observations in our sample to the corresponding expected income distribution, and which parametric form provides the best fit for SILC data. Following Cowell and Flachaire (2007) and Davidson and Flachaire (2007) we correct the Gini coefficient by replacing highest-income observations with values drawn from a parametric distribution and combining the corresponding parametric inequality measure for these incomes with a non-parametric measure for lower incomes. The following sections discuss the mechanics of fitting the alternative parametric forms to the data at hand.

2.2.1. Pareto Distribution

For the past century, the Pareto distribution has been applied to various socio-economic phenomena and is thought to be suitable to model the distribution of upper incomes. The Pareto distribution can be described by the following cumulative density function:

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

where α is a fixed parameter called the Pareto coefficient and x is the variable of interest (income in our case) and L is the lowest value allowed for x in the case of left censoring. The corresponding probability density function, allowing for right-censoring at H (separating potentially contaminated top-income observations, $H \leq x \leq \infty$, from reliable bottom observations, $0 \leq x \leq H$), is

$$f(x) = \frac{\alpha L^\alpha}{x^{\alpha+1}} / \left(1 - \left(\frac{L}{H}\right)^\alpha\right), \quad L \leq x < H \quad (5)$$

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

Parameter α in Equations (4) and (5) can be estimated using maximum likelihood from a right-truncated Pareto distribution, which also provides robust standard errors (Jenkins and Van Kerm 2015).

The Gini among the top k households can be derived from the expression of the corresponding Lorenz curve as follows

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

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 $4\alpha(\alpha - 1) / \left[\eta(\alpha - 2)(2\alpha - 1)^2(3\alpha - 2) \right]$ (Modarres and Gastwirth 2006), where η is the estimation sample size ($L \leq x < H$). The estimation error due to the potentially imprecise estimates of α is equal to $\epsilon / (2\alpha^2 - 2\alpha - 2\alpha\epsilon + \epsilon + 0.5)$, where ϵ is the standard error of $\hat{\alpha}$.

2.2.2. Generalized Beta Type 2 Distribution

Because the Pareto distribution is not representative of incomes in the middle or bottom of the income distribution, and because even among top incomes in some countries it may not follow the dispersion of incomes accurately, more flexible parametric distributions have been considered in recent literature. The four-parameter Generalized Beta distribution type 2 (GB2) has been suggested as providing better and more consistent fit for the distribution in various EU and US income surveys (Jenkins et al. 2011). It has the cumulative distribution function

$$F(x) = I\left(p, q, \frac{(x/b)^a}{1 + (x/b)^a}\right) \quad (7)$$

In this equation, $I(p, q, y)$ is the regularized incomplete beta function, in which the last argument, y , is income normalized to be in the unit interval. Parameters a , b , p , and q are parameters estimable with their standard errors by maximum likelihood. Because the right tail may be contaminated by top-income issues, right-truncation may be applied in the calculation of the GB2 density and model likelihood functions.

Moreover, as with the Pareto distribution, the GB2 distribution itself may not approximate well the bottom-most incomes, even though it tends to perform well in the middle and the top of the distribution. Jenkins et al. (2011, p. 69) propose left-truncating the distribution at the 30th percentile, a suggestion that this paper follows.⁶ Finally worth noting, the Gini under the estimated left- and right-truncated GB2 distribution could be computed by evaluating the corresponding generalized hypergeometric function ${}_3F_2(\hat{a}, \hat{b}, \hat{p}, \hat{q})$ (McDonald 1984; Jenkins 2009b).

2.3. Corrected Gini for EU States and EU-Wide

Replacing of observed top incomes with fixed Pareto or GB2 fitted values has the problem that it does not account for parameter-estimation error and sampling error in the available sample. The resulting Gini carries an artificially low standard error. An and Little (2007), and Jenkins et al. (2011) account for sampling error by drawing random values from the estimated distribution for all top incomes.

In the case of the EU SILC, we derive a corrected Gini coefficient across all EU member states as follows. The cumulative parametric distributions in Equations (4) and (7) are estimated at the level of each member state, and top incomes observed in each member state are replaced with random draws from the corresponding state-specific parametric distribution, as proposed by An and Little (2007), and Jenkins et al. (2011). Combining the observed lower-income values and the imputed top incomes across all EU member states allows us to derive a non-parametric estimate of the aggregate EU-wide Gini. Finally, repeating the exercise (bootstrapping) we obtain a *quasi-nonparametric* EU-wide Gini with its standard error (Reiter 2003).

As compared to the semi-parametric approach conventionally used in countries with homogeneous populations, this procedure allows the EU-wide distribution to include observations from both tails of state-level distributions, and preserve the original number of observations for each country. It also allows modalities such as custom truncation of state samples used for parametric estimation and for inequality measurement. Estimating the parametric distributions at the level of EU member states and replacing top incomes according to the estimated country-specific distributions

⁶ Indeed, during GB2 estimation on the SILC with Eurostat sampling weights, the algorithm could not converge due to the bottommost income observations (2.50 Euro/year or less). This indicates atypical distribution of the bottommost incomes. Indeed, there are over 100 observations in the SILC with annual income less than 100 Euro, suggesting measurement errors.

ensures that each state will have true lower incomes as well as replacement top incomes in the EU-wide data.⁷ The random draws of incomes ($x > H$) from the parametric distributions (estimated on incomes between L and H) can be combined with true lower incomes (up to H) as well as with incomes across EU states. Such flexible estimation of the EU-wide Gini and its standard error would generally not be possible with parametric estimates of the top-income Ginis.

Comparing the corrected state Ginis from the replacing analysis with the observed non-parametric Ginis would indicate whether the observed high incomes have been generated by Pareto- or GB2-like statistical processes, or whether the observed Gini is affected by top-income issues such as missing or non-representative values. A quasi-nonparametric Gini that is lower than the nonparametric Gini can be interpreted as evidence that some top incomes are extreme compared to those predicted under the parametric distribution. A higher quasi-nonparametric Gini would indicate that the observed top incomes are distributed more narrowly than would be predicted parametrically, potentially implying under-representation, censoring, or measurement errors in relation to high-income units in the sample.

An important decision in applying the replacing method relates to the range of incomes that should be replaced as potentially nonrepresentative or contaminated. Cowell and Flachaire (2007) choose a threshold at the 90th percentile of incomes. On the basis of the quality of fit in the United Kingdom income surveys, Jenkins (2017) advocates setting the threshold at top 1% or 5% incomes. We consider replacing between the top 1% and the top 10% of incomes with synthetic values contaminated only by randomness of the draw from the parametric distributions.

In conclusion, the reweighting and replacing methods differ in several respects and address different types of problems related to top incomes. Reweighting considers the entire income support and reweights all observations throughout the support according to the probability of non-response estimated with real data. Replacing keeps all observations up to the cut point unaltered while replacing all observations above the cut point with observations drawn from a theoretical distribution. Reweighting uses a probabilistic model drawing information from within and between regions' non-response rates to estimate the probability of non-response. Replacing does not make use of non-response rates or probabilistic models and uses instead estimated parameters from theoretical distributions to replace observations at the top. Reweighting is suited to address issues related to unit and item non-response and trimming whereas replacing is suited to address issues related to item underreporting, generic measurement errors, topcoding, and undue sensitivity of inequality measurement to the inclusion of rare extreme income observations.

3. Data

The methodologies outlined in the above section are evaluated using the set of national household surveys included in the 2011 round of the EU Statistics on Income and Living Conditions (SILC). This is a challenging set of surveys with different types of problems related to measurement issues that affect top incomes and inequality estimates.⁸

The SILC surveys, coordinated by a Directorate-General of the European Commission, Eurostat, cover 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. The data include relatively large sample sizes for each state but suffer from very different non-response rates across member states, and from limited potential for regional disaggregation. Average national non-response rates range from 3.3 to 50.7 percent across

⁷ Conversely, if all EU-wide incomes were used for estimation and replacement, this estimation and replacement would be largely done on the richest member states. Poorer states would then be represented with largely true incomes, while the richest states would be largely replaced, a dubious exercise. Moreover, while the Pareto law may hold for each EU member state, there is no guarantee that it would hold on incomes EU wide.

⁸ This analysis cannot be performed across multiple waves of SILC for several reasons: SILC was first collected only in 2004; Availability of countries has varied by wave; member states are not required to collect or publish sub-national non-response rates, and some statistical agencies have declined to compute them for the authors of this study citing lack of resources.

member states in the 2011 wave, and from 3.5 to 48.1 percent in 2009 (Tables S1 and S2 in the online Supplementary Material). These features allow for a limited number of model specifications to be used to reevaluate inequality under various measurement issues.⁹

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 SILC data are derived from country specific surveys which take different forms in different countries. However, in our case, they are an interesting set of data in that they are characterized by substantial diversity compared to other national surveys (Hlasny and Verme 2015). They are therefore a good benchmark to test how different top incomes correction methodologies perform under such diversity, provided that systematic cross-country differences are controlled for.¹⁰ One challenge is that incomes exhibit substantial cross-nation inequality, but relatively less inequality within nations, as evidenced by the difference between state-specific and EU-wide Gini indexes (refer to Tables S1 and S2). In fact, decomposition of the EU-wide Gini reveals that 67 percent of inequality arises solely from income differences between EU member states, and only 4 percent arises solely from within-state inequality, while 29 percent is due to an overlap of the between and within state inequality (2009 SILC shows analogous results).

With little overlap between income distributions in the richest and the poorest member states, when the reweighting correction method is run at the level of states (rather than within-state regions), it would effectively adjust the mass of entire member states in the calculation of the Gini. The vast majority of households in rich states would be assigned higher weights, and the majority of households in poor states would be assigned lower weights. This suggests that the analysis performed at a more geographically disaggregated level is warranted. To that end, we have collected unit non-response rates for NUTS-1 regions, that is geographic divisions, provinces or states of EU member countries.¹¹ Refer to Tables S2 and S3 in the online Supplementary Material. In what follows, we will primarily make use of the 2011 round of the SILC, and we will report on the 2009 round only on the margins. When not noted explicitly, the discussion refers to the 2011 round.

Household non-response rates (NRh) in SILC surveys are computed using Eurostat notation as:

$$NRh = 1 - \underbrace{\frac{\sum 1(db120 = 11)}{\sum 1(db120 \neq \emptyset) - \sum 1(db120 = 23)}}_{\text{Address contact rate}} \underbrace{\frac{\sum 1(db135 = 1)}{\sum 1(db130 \neq \emptyset)}}_{\text{Rate of complete interviews accepted}}, \quad (8)$$

where $1(\cdot)$ is a binary indicator function, db120 is the record of contact at the address, db130 is the household questionnaire result and db135 is the household interview acceptance result. Addresses that could not be located or accessed ($db120 \leq 22$) are accounted for in the address contact rate, while non-existing, non-residential, non-occupied and non-principal residence addresses ($db120 = 23$) are omitted. Rate of complete interviews accepted is the accepted interviews (i.e., at least one personal interview in household accepted) among all households completing, refusing to cooperate, temporarily absent, or unable to respond due to illness, incapacity, language or other problems.

Sampling weights available in SILC (db090) account for units' probability of selection, limited correction for the probability of non-response by different population subgroups, and calibration of sample representativeness vis-à-vis the distribution of households and persons in the target population, including by sex, age, household size and composition and NUTS-2 region (European Commission 2006).

⁹ For more information on the SILC see: <http://ec.europa.eu/eurostat/web/income-and-living-conditions>.

¹⁰ Sampling weights in the SILC are distributed very widely, from essentially zero to 38,357.27 (mean 901.89, standard deviation 1050.31) in the 2011 round. This also suggests that comparing unweighted, SILC weighted, and our non-response probability weighted statistics may yield very different estimates. Moreover, sampling weights in the 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.

¹¹ For Cyprus, Estonia, Germany, Iceland, Latvia, Lithuania, Luxembourg, Malta, Portugal, Slovenia and Switzerland, non-response rates are available by the degree of urbanization (db100 variable): dense, intermediate or thin level of population density. In 2009 for Slovakia and the UK, only nationwide non-response rates are available.

The income variable that is best comparable across SILC national surveys is the equivalized disposable income, hx090. The equivalized household size is computed as $hx050 = 1 + 0.5 \times (\text{adults} - 1) + 0.3 \times \text{children}$, where adults are those aged 14 or over at the end of the income reference period and children are those aged 13 or less.¹² Income is not adjusted for cost-of-living differences across EU member states for conceptual and empirical reasons. First, workers in the European Single Market can spend their income in any jurisdiction as well as on Internet purchases, circumventing local price differentials. Second, it is unclear which single cross-country price index should be applied to workers' earnings, consumption and savings, and the SILC database does not provide such a price index. The income aggregate across countries may also have a different capacity to capture capital income either by design or by practice.

Finally, many of the EU statistical agencies combine survey and administrative information such as tax and social security records to estimate income (refer to individual chapters of Jäntti et al. (2013)). This may result in a more accurate estimation of incomes as compared to countries that do not adopt this strategy. If this is the case, both the reweighting and replacing methods should show (correctly) a lower bias as for any survey with better quality data. However, these techniques vary across countries and can play a role when comparing estimated biases across countries. Considering the fact that the original survey instruments differ and that the income aggregates are not identical in their composition, estimations presented in this paper are not strictly comparable across countries. Moreover, the influence of each country in the overall estimation for the EU Gini is also affected by these factors.

4. Results

4.1. Reweighting

Table 1 presents the benchmark results for the *reweighting* correction method described in Equations (1)–(3). Equivalized disposable income is used as the outcome variable whose inequality is being measured, as well as the main element of x_{ij} (in logarithmic form). Binary indicators for European countries are also included as element of x_{ij} in light of the high heterogeneity in incomes, inequalities and non-response rates across Europe.¹³

The main finding is that households' survey response probability is related negatively to disposable income. The estimated coefficient on log income ($\hat{\theta}_2$) is negative and significantly different from zero, 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. The unweighted corrected Gini coefficient is 48.34. This is higher than the uncorrected and unweighted Gini by 3.25 percentage points, statistically highly significant. Making use of the sampling weights provided by national statistical agencies does not affect these findings. The correction for unit non-response in this case amounts to 3.70 percentage points of the Gini.¹⁴

¹² There are two editions of the EU-SILC survey produced by Eurostat. The Production Data Base (PDB) includes all available variables for responding and nonresponding households, while a Users Data Base (UDB) excludes nonresponding units and variables that could potentially allow identification of households. Related to our analysis, the PDB includes variables *DB120*, *DB130* and *DB135*, defining responding and non-responding households, *DB060-DB062*, identifying primary sampling units, and *DB075*, separating the traditional non-response rate (households interviewed for the first time) from the attrition rate (households from the 2nd to the 4th interview). Unfortunately, the PDB is not shared with users for confidentiality reasons, so in this study we rely on the UDB datasets.

¹³ This includes 27 country indicators, with Hungary and Slovakia; Denmark and Norway; and Ireland and Island respectively sharing single indicators due to their empirical similarities, and The Netherlands serving as a baseline country. Alternatively, 12 regional indicators plus a baseline were considered, in agreement with geopolitical division of Europe and with empirical distribution of incomes, inequalities and non-response rates across countries. Refer to Table S16 in the online Supplementary Material. However, this less parameterized specification still produced inconsistent results due to the remaining systematic heterogeneity within the 13 European regions.

¹⁴ Note that applying the sampling weights to the distribution of incomes uncorrected for unit non-response reduces the Gini in the SILC by 5.7 percentage points. This happens because sampling weights in the SILC (correcting for various sampling issues including region-level non-response) and the estimated non-response correction weights are related negatively for most households. SILC sampling weights are higher among households with atypical incomes, and lower among

To the extent that applying the statistical agency weights amounts to some double-correcting for non-response and these corrections interact with each other arbitrarily, we can estimate a quasi-difference-in-difference type of effect of weighting. The stand-alone correction for non-response is estimated at 3.60 percentage points of the Gini (48.34–45.10). The stand-alone correction for non-representative sampling is estimated at −6.19 percentage points of the Gini (38.91–45.10). Adding these effects to the uncorrected Gini, we conclude that the robust Gini is 42.15. This figure is slightly lower than the original estimate of 42.61, suggesting that the double-correction of non-response is responsible for a 0.46 percentage-point inflation of the Gini. In conclusion, *reweighting* is consistent in finding an upward correction of the Gini of between 3.25 and 3.70 percentage points.

Table 1. Benchmark results of Gini correction for unit non-response bias.

Variable	Coefficient Estimate
Intercept	12.377 (1.306)
Log(income)	−1.047 (0.127)
AT	−0.571 (0.156)
BE	−1.386 (0.134)
BG	−1.360 (0.414)
CH	−0.112 (0.164)
CY	0.146 (0.311)
CZ	−1.212 (0.227)
DE	0.042 (0.175)
EE	−2.221 (0.232)
EL	−1.611 (0.169)
ES	−0.381 (0.187)
FI	−0.248 (0.158)
FR	−0.452 (0.145)
HR	−3.035 (0.219)
IE, IS	−0.794 (0.155)
IT	−0.866 (0.133)
LT	−1.790 (0.289)
LU	−0.982 (0.144)
LV	−2.249 (0.251)
MT	−0.533 (0.294)
DK, NO	−1.289 (0.135)
PL	−1.583 (0.241)
PT	−0.259 (0.348)
RO	−0.869 (0.719)
SE	−1.229 (0.133)
SI	−1.284 (0.165)
HU, SK	−1.330 (0.265)
UK	−0.972 (0.141)
Regions j	31 member states
Households i	238,383
Uncorrected Gini	45.10 (0.08)
Gini using stat. agency weights	38.91 (0.13)
Gini corrected for unit non-response bias	48.34 (0.84)
Gini corrected for unit non-resp. bias, with sampling wts.	42.61 (0.83)
Unit non-response bias	3.25
Bias (using sampling wts.)	3.70

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

households in the middle of the national income distributions. Hence, combination of the two sets of weights serves to dampen the effect of inflating the representation of atypical units with very low incomes. This dampening—which lowers the estimate of inequality—overshadows the double-correction for unit non-response among top-income households.

Using the results in Table 1 and the estimated non-response correction weights, we can re-estimate the Ginis for each EU member state (Table 2, last column). The corrected Gini increases by 0.2–6.5 percentage points, with the exception of Belgium and Slovakia (20.0 and 9.3 pc.pt. correction, respectively). The corrected Ginis for Belgium and Slovakia carry high standard errors and should be viewed with great caution.¹⁵ Across the 29 EU member states (excluding the two outliers, and without accounting for states' population or sample sizes), the estimated Gini correction is strongly positively associated with states' mean income (correl. +0.541), mean non-response rate (correl. +0.219) and the count of regions used for sub-national disaggregation (correl. +0.488).¹⁶ Finally, refer to the discussion on the survey instruments, income aggregates and combination with administrative data to understand other potential sources of cross-country differences in estimated biases.

Table 2. Non-response rate and income distribution by member state, 2011 SILC.

Member State	Sub-National Regions	Househds.	National Non-Response Rate (%)	Mean Equivalized Disposable Income (Euro)	State Gini, SILC Weighted Households	Pure within-Region Contrib. (%)	Pure between-Region Contrib. (%)	State Gini, SILC Weighted & Non-Response Corrected
Austria	3	6183	22.6	23,713.37	27.59 (0.40)	34.2	10.4	29.54 (0.75)
Belgium	3	5897	36.7	21,622.14	27.63 (0.91)	39.5	10.7	47.61 (18.45)
Bulgaria	2	6548	7.5	3415.42	35.99 (0.58)	49.3	14.5	37.88 (1.03)
Croatia	1	6403	43.3	5981.46	32.07 (0.36)	–	–	32.81 (0.57)
Cyprus	3	3916	10.2	20,084.84	31.65 (1.02)	44.3	16.8	36.41 (4.35)
Czech Rep.	8	8865	17.1	8402.77	25.91 (0.37)	12.4	20.7	27.53 (0.57)
Denmark	1	5306	44.4	28,441.21	27.45 (0.55)	–	–	31.00 (1.30)
Estonia	2	4980	26.0	6475.47	32.62 (0.55)	54.4	12.3	34.15 (0.82)
Finland	4	9342	18.1	23,870.09	26.83 (0.37)	24.7	20.3	29.71 (1.92)
France	21	11,348	18.0	24,027.78	30.84 (0.45)	7.2	20.3	36.99 (1.72)
Germany	3	13,473	12.6	21,496.55	30.21 (0.33)	41.0	7.1	32.41 (0.77)
Greece	4	5969	26.5	12,704.72	32.92 (0.57)	27.6	17.0	35.67 (1.10)
Hungary	3	11,680	11.2	5146.29	26.86 (0.26)	34.1	22.0	27.58 (0.31)
Iceland	2	3008	24.8	20,668.26	24.99 (0.64)	53.8	6.0	28.00 (1.68)
Ireland	8	4333	19.6	39,831.65	32.92 (0.56)	14.9	23.8	34.82 (1.10)
Italy	5	19,234	25.0	18,353.37	31.72 (0.29)	21.6	23.7	35.56 (1.00)
Latvia	2	6549	18.9	5048.72	34.98 (0.39)	49.0	17.0	36.46 (0.48)
Lithuania	2	5157	18.6	4588.81	33.02 (0.57)	50.0	16.6	33.95 (0.65)
Luxembourg	3	5442	43.3	37,232.63	27.32 (0.47)	35.5	12.7	29.42 (0.86)
Malta	2	4070	11.8	12,167.55	28.29 (0.44)	81.5	1.9	28.95 (0.52)
The Netherlands	1	10,469	14.5	22,726.06	25.66 (0.34)	–	–	27.01 (0.56)
Norway	1	4621	50.7	38,616.14	24.98 (0.59)	–	–	29.39 (3.05)
Poland	6	12,861	14.9	5849.61	32.10 (0.39)	17.5	10.1	34.32 (0.73)
Portugal	3	5740	7.9	10,462.34	35.07 (0.57)	32.8	19.2	36.35 (0.72)
Romania	4	7614	3.3	2447.42	32.37 (0.39)	25.0	13.3	32.58 (0.41)
Slovakia	4	5200	14.5	6983.48	27.30 (1.26)	28.5	15.0	36.58 (9.42)
Slovenia	1	9246	23.8	12,714.07	25.84 (0.29)	–	–	26.54 (0.38)
Spain	19	12,900	37.2	14,584.40	32.67 (0.26)	6.7	23.6	33.03 (0.29)
Sweden	1	6694	36.5	23,727.45	25.76 (0.36)	36.8	9.0	28.65 (2.52)
Switzerland	3	7502	24.0	39,327.92	30.28 (0.49)	42.6	12.0	34.82 (1.60)
UK	37	8009	27.3	20,843.59	32.85 (0.57)	3.1	24.5	39.32 (2.88)
Wtd. Mean [EU wide]	5.23 [162]	7695 [238],[559]	23.9	17,929.58	29.61 [38.91]	–	–	32.99 [42.61]

Note: Non-response rate is reported in the member-states' Intermediate/Final Quality Reports at the state level as *NRh* for total sample. 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. Source: EU-SILC data in World Bank database; Ireland data from Luxembourg Income Study database.

¹⁵ The high corrections of the Ginis in Belgium and Slovakia are not due to atypical distributions of incomes across national regions—Gini decomposition shows similar within- and between-region components (Table 2, two columns before the last column). Instead, it is due to exceptionally thin top-income distributions with rare extreme incomes. Tables S4–S9 show that the Pareto coefficients estimated among the highest quartile of incomes in Belgium (particularly from the 75–80th percentile to the 92–94th percentile) are the highest or among the highest of all EU member states. Pareto coefficients estimated for Slovakia are also above average, but not exceptionally high. These thin top ends of the income distribution suggest that the few observed extreme incomes, when reweighted, can have great influence on the measurement of inequality. This also explains the high standard errors on the Ginis.

¹⁶ The number of regions *j* selected for the estimation of Equation (3) determines the weight that the model attributes to *within*-region as opposed to *between*-regions information and this choice leads to significantly different estimations of the correction bias. Analyses using finer degrees of disaggregation have been found to typically yield lower corrections for unit non-response (Hlasny 2016; Hlasny and Verme 2017, 2018). In Tables 1 and 2, however, the estimates come from a model on the entire set of 31 member states, using a fixed degree of disaggregation into 162 regions.

4.2. Replacing

Next, we use a methodology first proposed by Cowell and Victoria-Feser (2007) to test the sensitivity of the Gini coefficients to extreme or non-representative observations on the right-hand side of the distribution. We correct for the influence of potentially contaminated top incomes using an estimated Pareto or generalized beta distribution as discussed in the methodological section. The analysis is performed at the level of individual EU member states, so that the replaced income values would come from all states rather than just from a handful of the richest states. Table 3 presents quasi-nonparametric estimates of the Gini coefficients obtained by replacing the highest top 1–10 percent of income observations in each state with values imputed from the estimated Pareto distribution left-truncated at the 75–85th percentile of incomes and right-truncated at the 92–99th percentile of incomes. (Tables S4–S9 in the online Supplementary Material show the results for each member state.) Lower right-truncation, such as at the 90th percentile, could not be performed because it would leave small national sample sizes for estimation (say, 85–90th percentile incomes), particularly compared to the range of incomes for replacing (say, 91st percentile incomes and above), and would yield volatile or excessively high Ginis. Recall that the estimation is performed at the national level, and national samples are not large (Table 2). By the same token, lower left-truncation would compromise the quality of fit of the Pareto distribution.

The choice of right truncation is a critical parameter because it affects which observations will be classified as uncontaminated and will be used to estimate the parametric distribution, and which observations labeled as suspect will be replaced with values drawn from the distribution. The corrected inequality index will be based on the actual income observations to the left of the right-truncation point, and only on synthetic values to the right of that point. Since there is no theoretically favored point for left- or right-truncation, and there is limited empirical guidance on how to set them particularly in a new dataset for a group of countries such as the EU-SILC, we consider a range of cutoff points. Values of the estimated parameters, measures of model fit, and the estimated corrections for the Gini can be used to determine which ranges of incomes are best suited for estimation and for out-of-sample prediction.

Results are shown in Table 3. The table has three sets of rows, for left-truncation set at the 85th, 80th and 75th percentile of national incomes. We find that the choice over left truncation in estimation does not affect the measurement of inequality significantly. The Ginis are corrected by -0.2 to $+4.4$ percentage points regardless of the left-truncation point. On the other hand, right truncation affects the measurement systematically. This should not be surprising, because right-truncation in this exercise affects not only the estimation of the Pareto distribution, but also the extent of replacing observed top incomes with values drawn from the national parametric distributions. When only 1% of top incomes are replaced, the Gini typically falls by 0.02 to 0.20 points, suggesting that the observed topmost incomes are extreme and over-represent the incomes of the richest 1 percent in the population as predicted by the estimated national Pareto distributions.¹⁷ However, when 5–8 percent of observed top incomes are replaced, the Gini rises by 0.39 to 4.36 percentage points, suggesting that in this group (and particularly in the second ventile of the national distributions) the observed incomes typically underrepresent the incomes in the population due to unit non-response and other biases. These latter results are consistent with the results provided by reweighting potentially suggesting that the Pareto distribution mimics rather well the top decile of the real income distribution but not the very top of the distribution (top 1 or up to top 5 percent).

¹⁷ Analogous replacement was also done for the top 0.2, 0.5 and 0.7 percent of incomes. The effects of these replacements are smaller than those in Table 3, as they reflect the replacement of individual outlying observations.

Table 3. Correction by replacing incomes with random draws from national Pareto distributions.

Correction of Extreme Observations	Sampling Correction	Sample Size η k obs. Replaced	Gini	Bias in Original Gini (pc.pt.)
Estimation on top 15– h th percentile of incomes				
Semi-param. estimation, $h = 1\%$	Unweighted	$\eta = 33,380$ $k = 2400$	44.92 (0.07)	−0.18
	Eurostat weights	$\eta = 34,475$ $k = 2587$	38.71 (0.13)	−0.20
Semi-param. estimation, $h = 5\%$ ⁱ	Unweighted	$\eta = 23,841$ $k = 11,939$	45.49 (0.11)	+0.39
	Eurostat weights	$\eta = 24,517$ $k = 12,545$	38.85 (0.14)	−0.06
Semi-param. estimation, $h = 6\%$ ⁱ	Unweighted	$\eta = 21,463$ $k = 14,317$	45.87 (0.16)	+0.77
	Eurostat weights	$\eta = 21,994$ $k = 15,068$	43.27 (11.01)	+4.36
Estimation on top 20– h th percentile of incomes				
Semi-param. estimation, $h = 1\%$	Unweighted	$\eta = 45,295$ $k = 2400$	44.98 (0.07)	−0.12
	Eurostat weights	$\eta = 46,702$ $k = 2587$	38.81 (0.13)	−0.10
Semi-param. estimation, $h = 5\%$	Unweighted	$\eta = 35,756$ $k = 11,939$	45.72 (0.10)	+0.62
	Eurostat weights	$\eta = 36,744$ $k = 12,545$	39.66 (0.16)	+0.75
Semi-param. estimation, $h = 8\%$	Unweighted	$\eta = 23,860$ $k = 19,086$	47.26 (0.18)	+2.16
	Eurostat weights	$\eta = 29,302$ $k = 19,987$	42.15 (0.47)	+3.24
Estimation on top 25– h th percentile of incomes				
Semi-param. estimation, $h = 1\%$	Unweighted	$\eta = 57,218$ $k = 2400$	45.04 (0.08)	−0.06
	Eurostat weights	$\eta = 58,841$ $k = 2587$	38.89 (0.14)	−0.02
Semi-param. estimation, $h = 5\%$	Unweighted	$\eta = 47,679$ $k = 11,939$	46.09 (0.17)	+0.99
	Eurostat weights	$\eta = 48,883$ $k = 12,545$	40.12 (0.19)	+1.21
Semi-param. estimation, $h = 8\%$	Unweighted	$\eta = 40,532$ $k = 19,086$	47.89 (0.20)	+2.79
	Eurostat weights	$\eta = 41,441$ $k = 19,987$	42.41 (0.46)	+3.50

Notes: Pareto coefficients are estimated on non-contaminated income observations (sample size η ; $L \leq x < H$; H is income corresponding to the 100– h th percentile) using maximum likelihood, and are then used to impute values for the k top-income observations. Parametric replacement is done at the national level. Europe-wide Ginis and their standard errors are computed across all national quasi-nonparametric income distributions, and are bootstrapped. For clarity, Ginis and their standard errors are multiplied by 100. Sampling weights are adopted from Eurostat. ⁱ Right-truncation here is higher than in the models below. Any lower right-truncation point than this leads to overly large and erratic Gini estimates due to small national estimation samples (i.e., range of income quantiles on which Pareto distribution is fit) and comparatively large national prediction samples (i.e., quantiles for which Pareto estimates are drawn). Refer to Table S5.

Tables S4–S9 in the online Supplementary Material present the Pareto coefficients α and semiparametric Ginis estimated for each EU member state.¹⁸ Like under the reweighting approach,

¹⁸ The parametric Gini estimates among top incomes in Tables S4–S9 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 numbers drawn randomly from the corresponding Pareto distributions, and bootstrapping the exercise.

the corrections of the Ginis across individual states are in line with the EU-wide corrections with some notable exceptions. Estimated Pareto coefficients are low in several states—notably Cyprus, Estonia, Ireland and Latvia—on account of a narrow dispersion of top incomes and rare extreme incomes, leading to high corrected Ginis in those states. On the other end of the spectrum, Belgium, Iceland, Norway, Slovenia and Sweden have high estimated Pareto coefficients leading to lower corrected Ginis. The effects of top-income replacement on the Ginis are dampened by the fact that replacement is applied only to top incomes, while original values are used for the rest of incomes. In comparison, the reweighting method affected the contribution of all income observations, leading to even larger corrections of the Gini.

The variation in the Pareto coefficients across model specifications indicates that the estimated α depends systematically on the way income observations are weighted, and on the range of top incomes under analysis. Pareto coefficients are estimated somewhat higher in the income distribution weighted by Eurostat sampling weights than in the unweighted distribution. Moreover, the higher the values of incomes evaluated in terms of the left and right truncation points, the higher the Pareto coefficient, and thus the lower the corresponding inverted Pareto coefficient β , the estimated top income share and the Gini. The highest Pareto coefficients are obtained when the national distributions are left-truncated at the 85th percentile and right-truncated at the 99th percentile. That suggests that extreme income dispersion may be a problem among the topmost 1% of incomes and between the 75th and 85th percentile, but not as much between the 85th and 99th percentiles.

One potential criticism of the Pareto distribution is that it relies on only one parameter to fit true top incomes. The fit of the one-parameter Pareto distribution to European and other income distributions has been questioned (Jagielski and Kutner 2013; Jenkins 2017). In the following paragraphs we re-estimate the semi-parametric Gini coefficients assuming top incomes to be distributed as under the generalized beta distribution. To do this, we estimate the generalized beta distribution that provides the best fit for the distribution of top 70 percent of incomes in each state, and then use predicted values to compute a parametric Gini coefficient for the state. To derive an EU-wide Gini, we use values drawn randomly from the parametric distributions to replace topmost incomes in each state, and combine these replacement top-income values with actual lower incomes to derive the Gini quasi-nonparametrically.¹⁹

Table 4 reports the main results for the EU at large, and Tables S10–S15 in the online Supplementary Material report model coefficients and parametric Ginis for individual EU member states. Comparing the Ginis in Table 4 to the nonparametric estimates in Table 1, we find that the quasi-nonparametric Ginis under the assumed generalized beta distribution are systematically lower, implying that actual incomes may be distributed more unequally than incomes predicted under that distribution. The downward correction of the Gini is up to 3.3 percentage points and 1.4 percentage points on average across the 6 model specifications reported.

Compared to the Pareto distribution, the corrections to the Gini coefficients under the generalized beta distribution are consistently negative, but of a similar magnitude in absolute value. This indicates that the estimated generalized beta distributions predict a narrower dispersion of top incomes than the estimated Pareto distributions, but both estimations give rise to concerns about top-income biases of a similar magnitude, 0–4 percentage points of the Gini.

These Ginis from random draws are very similar to the smooth-distribution Ginis in Tables S4–S9, but have slightly higher standard errors due to sampling errors.

¹⁹ To validate the procedure, we again compare the parametric and quasi-nonparametric Ginis in each state (refer to the previous footnote). Indeed, 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.

Table 4. Correction by replacing incomes with random draws from national GB2 distributions.

Correction of Extreme Observations	Sampling Correction	Sample Size η k obs. Replaced	Gini	Bias in Original Gini (pc.pt.)
Estimation on top 70– h th percentile of incomes				
Semi-param. estimation, $h = 1\%$	Unweighted	$\eta = 164,423$ $k = 2400$	43.89 (0.05)	−1.21
	Eurostat weights	$\eta = 167,932$ $k = 2587$	37.64 (0.08)	−1.27
Semi-param. estimation, $h = 5\%$	Unweighted	$\eta = 154,944$ $k = 11,939$	44.86 (0.08)	−0.24
	Eurostat weights	$\eta = 158,093$ $k = 12,545$	36.80 (0.09)	−2.11
Semi-param. estimation, $h = 10\%$	Unweighted	$\eta = 143,233$ $k = 14,317$	44.73 (0.14)	−0.37
	Eurostat weights	$\eta = 145,699$ $k = 15,068$	35.57 (0.07)	−3.34

Notes: GB2 coefficients are estimated on non-contaminated income observations (sample size η ; $L \leq x < H$; L is income corresponding to the 30th percentile; H is income corresponding to the 100- h th percentile) using maximum likelihood. Quasi-nonparametric Ginis and their standard errors are bootstrap estimates, and are multiplied by 100.

Coefficient estimates presented in Tables S10–S15 carry for the most part acceptable standard errors and are rather consistent across model specifications with different sampling weights and right-truncation points. There are unclear patterns in the estimated coefficients between the analyses performed under alternative weighting schemes (unweighted versus Eurostat weighted) and alternative sample cutoff points (90th, 95th, 99th percentile). The higher the range of incomes included in estimation (up to the 95th or the 99th percentile), the systematically lower the distributional shape parameter a , but the other shape parameters (p , q) and the scale parameter (b) vary non-systematically. As a byproduct of our analysis, we can confirm that the generalized beta distribution cannot be easily approximated by Singh-Maddala or Dagum distributions as \hat{p} and \hat{q} , respectively, are significantly different from unity across most EU member states, under all weighting schemes and sample-truncation points in the analysis.

The estimated parametric Ginis vary greatly across EU member states, due to heterogeneous distributions of incomes and sampling weights across states, different sample sizes, and different quality of fit of the parametric GB2 distributions. Like in the case of reweighting and Pareto-replacing estimation, several states end up with outlying parametric estimates of their Ginis subject to high standard errors. Across multiple runs of the analysis (in Tables S10–S15), Belgium, Bulgaria, Finland, Greece, Ireland, Latvia, Norway and Slovenia end up with unreasonably high parametric estimates of their Ginis, while Denmark, Germany, Iceland, Slovakia and Sweden end up with unreasonably low Ginis.

5. Discussion

This study has evaluated two methods—*reweighting* and *replacing*—for correcting top-income biases generated by known data issues including unit and item non-response and more generally representativeness issues of top-income observations. The joint use of two distinct statistical methods for correcting top-income biases, sensitivity analysis of their technical specifications, and analysis of their performance on a challenging heterogeneous household survey were methodological contributions of this study.

Using the reweighting approach and the 2011 wave of the SILC, the paper finds a significant 3.3–3.7 percentage point downward bias in the Gini index.²⁰ The weighted Europe-wide Gini index is estimated at 42.61 percent as compared to a non-corrected Gini of 38.91 percent. The average Gini for the 31 European countries considered is estimated at 32.99 percent as compared to an uncorrected Gini of 29.61 percent.

Similar results are found using the replacing method with the Pareto distribution but only when the cutoff point for replacing is below 95 percent. The use of higher cutoff points yields very low biases, below 1 percentage point of the Gini. Given that top-income biases are expected to be higher at the very top, it is possible that the Pareto distribution does not mimic well the European income distribution at the very top. This may be due to the limited flexibility offered by the one-parameter Pareto distribution.

Repeating the replacing exercise with the four-parameter GB2 distribution does not improve our findings. Our estimates of inequality fall by 0.2–3.3 percentage points of the Europe-wide Gini, while the Ginis for individual member states are estimated very widely and often unreasonably low or high. We conclude that the popular 1–4 parameter distributions such as the Pareto and the GB2 distributions are not well suited to model the topmost incomes across a heterogeneous sample of distributions, and that alternative distributions should be sought to model the very top ends. The fact that these distributions were proposed and initially tested in the 20th century combined with the sharp growth of incomes at the very top of the distribution in the 21st century in Europe and elsewhere may contribute to explain this shortcoming.

Another problem with the replacing methods, similarly to the traditional treatments for item nonresponse, is that they rely on an assumption that other income observations are valid and accurate. Replacing methods assume away measurement issues below the cutoff point. At the same time, the parametric distributions proposed yield a wide range of empirical results (in Tables 3 and 4), indicating that parameters calibrated with the lower parts of the income distributions do not offer insights of any accuracy about the very top.

In perspective of the findings from the reweighting and parametric replacing exercises, we also conclude that the systematic under-representation of top-income households due to unit nonresponse is a more worrying problem than other potential contaminations of the top-income distribution for inequality measurement. Unit non-response leads to a systematic downward bias in the measurement of the Gini coefficient by 3–4 percentage points, while the balance of other top-income biases remains unclear, and has been estimated in this study widely at between a –3 and a +4 percentage point adjustment to the Gini.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2225-1146/6/2/30/s1>. Tables S1–S3 show summary statistics and unit nonresponse rates in national surveys. Tables S4–S9 show additional results of Pareto replacement for individual EU member states. Tables S10–S15 show additional results of replacement using Generalized Beta II distribution also for individual EU member states. Table S16 shows the considered delineation of 13 European regions.

Author Contributions: Both authors contributed equally to the paper.

Funding: This research received no external funding.

Acknowledgments: The authors are grateful to Ragui Assaad and Roy Van der Weide for peer reviewing an earlier draft of the paper, and to participants at the ECINEQ conference (Luxembourg 2015) for helpful discussion. All remaining errors are ours.

Conflicts of Interest: The authors declare no conflict of interest.

²⁰ This analysis cannot be performed across multiple waves of SILC for several reasons: SILC was first collected only in 2004; Availability of countries has varied by wave; member states are not required to collect or publish sub-national non-response rates, and some statistical agencies have declined to compute them for the authors of this study citing lack of resources.

References

- Alvaredo, Facundo, and Thomas Piketty. 2014. Measuring top incomes and inequality in the Middle East: Data limitations and illustration with the case of Egypt. Working Paper 832, Economic Research Forum, Giza, Egypt.
- Alvaredo, Facundo, Lydia Assouad, and Thomas Piketty. 2017. Measuring Inequality in the Middle East 1990–2016: The World's Most Unequal Region? WID. Available online: <http://wid.world/document/alvaredoassouadpiketty-middleeast-widworldwp201715/> (accessed on 29 May 2018).
- An, Di, and Roderick J. A. Little. 2007. Multiple imputation: An alternative to top coding for statistical disclosure control. *Journal of the Royal Statistical Society A* 170: 923–40. [CrossRef]
- Atkinson, Anthony Barnes, and John Micklewright. 1983. On the reliability of income data in the family expenditure survey 1970–1977. *Journal of the Royal Statistical Society Series A* 146: 33–61. [CrossRef]
- Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. Top incomes in the long run of history. *Journal of Economic Literature* 49: 3–71. [CrossRef]
- Bartels, Charlotte, and Maria Metzger. 2017. An Integrated Approach for Top-Corrected Ginis. IZA DP #10573. Available online: <https://www.iza.org/publications/dp/10573/an-integrated-approach-for-top-corrected-ginis> (accessed on 29 May 2018).
- Brachmann, Klaus, Andreas Stich, and Mark Trede. 1996. Evaluating parametric income distribution models. *Allgemeines Statistisches Archiv* 80: 285–98.
- Brzezinski, Michał. 2013. Parametric modelling of income distribution in Central and Eastern Europe. *Central European Journal of Economic Modelling and Econometrics* 3: 207–30.
- Burkhauser, Richard V., Shuaizhang Feng, and Jeff Larrimore. 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. [CrossRef]
- Burricand, Carine. 2013. Transition from survey data to registers in the French SILC survey. In *The Use of Registers in the Context of EU-SILC: Challenges and Opportunities*. Edited by Markus Jäntti, Veli-Matti Törmälehto and Eric Marlier. Luxembourg: European Union.
- Butler, Richard J., and James B. McDonald. 1989. Using incomplete moments to measure inequality. *Journal of Econometrics* 42: 109–19. [CrossRef]
- Cowell, Frank A., and Emmanuel Flachaire. 2007. Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics* 141: 1044–72. [CrossRef]
- Cowell, Frank A., and Maria-Pia Victoria-Feser. 1996. Poverty measurement with contaminated data: A robust approach. *European Economic Review* 40: 1761–71. [CrossRef]
- Cowell, Frank, and Maria-Pia Victoria-Feser. 2007. Robust Lorenz curves: A semiparametric approach. *Journal of Economic Inequality* 5: 21–35. [CrossRef]
- Dagum, Camilo. 1980. The generation and distribution of income, the Lorenz curve and the Gini ratio. *Economie Appliquée* 33: 327–67.
- Davidson, Russell, and Emmanuel Flachaire. 2007. Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics* 141: 141–66. [CrossRef]
- Deaton, Angus. 2005. Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and Statistics* 87: 20–22. [CrossRef]
- European Commission. 2006. *EU-SILC User Database Description*. EU-SILC/BB D(2005). Luxembourg: European Commission.
- Frick, Joachim R., and Kristina Krell. 2010. Measuring Income in Household Panel Surveys for Germany: A Comparison of EU-SILC and SOEP. SOEP paper 265, German Institute for Economic Research (DIW), Berlin, Germany.
- Hlasny, Vladimir. 2016. Unit Nonresponse Bias to Inequality Measurement: Worldwide Analysis Using Luxembourg Income Study Database. LIS Technical Paper, Luxembourg Income Study, Luxembourg City, Luxembourg.
- Hlasny, Vladimir, and Vito Intini. 2015. Representativeness of Top Expenditures in Arab Region Household Surveys. UN-ESCWA/EDID Working Paper #11, United Nations Beirut Economic and Social Commission for Western Asia, Beirut, Lebanon.
- Hlasny, Vladimir, and Paolo Verme. 2015. Top Incomes and the Measurement of Inequality: A Comparative Analysis of Correction Methods Using Egyptian, EU and US Survey Data. ECINEQ Conference Paper 2015-145, Society for the Study of Economic Inequality, Verona, Italy.

- Hlasny, Vladimir, and Paolo Verme. 2017. The Impact of Top Incomes Biases on the Measurement of Inequality in the United States. ECINEQ Working Paper #2017-452, Society for the Study of Economic Inequality, Verona, Italy.
- Hlasny, Vladimir, and Paolo Verme. 2018. Top incomes and the measurement of inequality in Egypt. *The World Bank Economic Review* 32: 428–55.
- Jagielski, Maciej, and Ryszard Kutner. 2013. Modelling of income distribution in the European Union with the Fokker–Planck equation. *Physica A: Statistical Mechanics and Its Applications* 392: 2130–38. [CrossRef]
- Jäntti, Markus, Veli-Matti Törmälehto, and Eric Marlier, eds. 2013. *The Use of Registers in the Context of EU–SILC: Challenges and Opportunities*, Eurostat, European Commission. Luxembourg: Publications Office of the European Union. Available online: <http://ec.europa.eu/eurostat/documents/3888793/5856365/KS-TC-13-004-EN.PDF> (accessed on 29 May 2018).
- Jenkins, Stephen P. 2007. Inequality and the GB2 Income Distribution. IZA Discussion Papers 2831, Institute for the Study of Labor (IZA), Bonn, Germany.
- Jenkins, Stephen P. 2009a. Distributionally-sensitive inequality indices and the GB2 income distribution. *Review of Income and Wealth* 55: 392–98. [CrossRef]
- Jenkins, Stephen P. 2009b. *GB2LFIT: Stata Module to Fit a GB2 Distribution to Unit Record Data*. Colchester: Institute for Social and Economic Research, University of Essex.
- Jenkins, Stephen P. 2017. Pareto models, top incomes and recent trends in UK income inequality. *Economica* 84: 261–89. [CrossRef]
- Jenkins, Stephen P., and Philippe Van Kerm. 2015. Paretofit: Stata Module to Fit a Type 1 Pareto Distribution. April 2007. Available online: <https://ideas.repec.org/c/boc/bocode/s456832.html> (accessed on 29 May 2018).
- Jenkins, Stephen P., Richard V. Burkhauser, Shuaizhang Feng, and Jeff Larrimore. 2011. Measuring inequality using censored data: A multiple-imputation approach to estimation and inference. *Journal of the Royal Statistical Society* 174: 63–81. [CrossRef]
- Korinek, Anton, Johan A. Mistiaen, and Martin Ravallion. 2006. Survey nonresponse and the distribution of income. *Journal of Economic Inequality* 4: 33–55. [CrossRef]
- Korinek, Anton, Johan A. Mistiaen, and Martin Ravallion. 2007. An econometric method of correcting for unit nonresponse bias in surveys. *Journal of Econometrics* 136: 213–35. [CrossRef]
- Lakner, Christoph, and Branko Milanovic. 2013. Global Income Distribution from the Fall of the Berlin Wall to the Great Recession. World Bank Policy Research Working Paper Series #6719, World Bank, Washington, DC, USA.
- Litchfield, Julie A. 1999. Inequality: Methods and Tools. Article for World Bank's Web Site on Inequality, Poverty, and Socio-Economic Performance. Available online: <http://siteresources.worldbank.org/INTPGI/Resources/Inequality/litchfie.pdf> (accessed on 29 May 2018).
- McDonald, James B. 1984. Some generalized functions for the size distribution of income. *Econometrica* 52: 647–63. [CrossRef]
- Mistiaen, Johan A., and Martin Ravallion. 2003. Survey Compliance and the Distribution of Income. Policy Research Working Paper #2956, The World Bank, Washington, DC, USA.
- Modarres, Reza, and Joseph L. Gastwirth. 2006. A cautionary note on estimating the standard error of the Gini index of inequality. *Oxford Bulletin of Economics and Statistics* 68: 385–90. [CrossRef]
- Pareto, Vifredo. 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. Available online: <https://www.cairn.info/ecrits-sur-la-courbe-de-la-repartition-de-la-riche--9782600040211.htm> (accessed on 29 May 2018).
- Reiter, Jerome P. 2003. Inference for partially synthetic, public use microdata sets. *Survey Methodology* 29: 181–88.
- Singh, S. K., and Gary S. Maddala. 1976. A function for the size distribution of income. *Econometrica* 44: 963–70. [CrossRef]
- Törmälehto, V.-M. 2017. High Income and Affluence: Evidence from the European Union Statistics on Income and Living Conditions (EU-SILC). Statistical Working Papers, Eurostat, Publications Office of the European Union, Luxembourg.

