BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Mind the wealth gap: a new allocation method to match micro and macro statistics on household wealth

by Michele Cantarella, Andrea Neri and Maria Giovanna Ranalli

October 2021

Number 646

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Mind the wealth gap: a new allocation method to match micro and macro statistics on household wealth

by Michele Cantarella, Andrea Neri and Maria Giovanna Ranalli

Number 646 – October 2021

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6627 (print)
ISSN 1972-6643 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

MIND THE WEALTH GAP: A NEW ALLOCATION METHOD TO MATCH MICRO AND MACRO STATISTICS ON HOUSEHOLD WEALTH

by Michele Cantarella[*], Andrea Neri[†] and Maria Giovanna Ranalli[‡]

Abstract

The financial and economic crisis that have shaken many countries in the last years have increased demand for timely, coherent and consistent distributional information for the household sector. In the Euro area, most of the national central banks collect such information through income and wealth surveys, which are often used to inform their decisions. These surveys, however, may be affected by non-response and under-reporting, determining a mismatch with macroeconomic figures from national accounts. In this paper, we develop a novel allocation method that extends proportional allocation and combines information from a power law (Pareto) model with imputation procedures based on calibration to address these issues simultaneously, when only limited external information is available. Finally, we produce distributional indicators for four Euro-Area countries, which are consistent with their national accounts.

JEL Classification: D31, E01, E21, N3.
Keywords: wealth distribution, non-response, measurement error, Pareto distribution, survey calibration, household finance and consumption survey.
DOI: 10.32057/0.QEF.2021/0646

Contents

[*] Centre for Consumer Society Research, University of Helsinki
[†] Economic and Financial Statistics Department, Banca d'Italia
[‡] Department of Political Science, Università degli Studi di Perugia

1 Introduction

The financial and economic crisis that has recently shaken many countries has increased demand for timely, coherent, and consistent distributional information on household income and wealth. Such information is featuring high on the agendas of national central banks (NCBs), which use it in several ways (Eurosystem Household Finance and Consumption Network, 2009). Distributional information is used for financial stability purposes, for example, to assess how much debt is concentrated in the hands of financially vulnerable households (see, for instance, Ampudia et al., 2016; Michelangeli and Rampazzi, 2016). Moreover, distributional information can be used to estimate the aggregate consumption response to wealth shocks when individual responses are heterogeneous (Christelis et al., 2021; Arrondel et al., 2019; Paiella, 2007; Guiso et al., 2005) and, more generally, to understand the interplay between monetary policy measures, especially non-standard ones, and the distribution of income and wealth (Mumtaz and Theophilopoulou, 2020; Colciago et al., 2019; Casiraghi et al., 2018; Coibion et al., 2017). Recent years have also been characterized by a surge of interest in the study of the dynamics of wealth accumulation over the last century (Frémeaux and Leturcq, 2020; Garbinti et al., 2018; Alvaredo et al., 2018).

Sample surveys are the main source of distributional information on household wealth. In the Euro area, most NCBs conduct the Eurosystem Household Finance and Consumption Survey (HFCS), which collects harmonized household-level data on households' finances and consumption (Eurosystem Household Finance and Consumption Network, 2009). The second source of information on household wealth comes from national accounts which record a country's stock of assets (both financial and non-financial) and liabilities at a particular point in time.

In theory, since the HFCS is designed to be representative of all households, aggregating this microdata should correspond to the macro aggregates. In practice, however, there are significant differences: aggregate totals based on surveys are often substantially below those found in national accounts. Before using the distributional information from survey data, it is, therefore, crucial to explain and possibly eliminate the differences between the two sources of information.[1]

There are several possible reasons for the differences (Expert Group on Linking Macro and Micro data, 2020). On the survey side, two relevant issues are unit non-response and measurement errors. There is substantial evidence that a household's decision on whether or not to participate in the survey is not at random. In particular, the wealthier the household, the more difficult it becomes to contact and to convince it to take part in the survey, to the point that, after an upper bound of wealth is reached, the probability of sampling reaches zero (Chakraborty et al., 2019; Kennickell, 2008, 2019; Vermeulen, 2018). Since these households own a large share of total wealth, their under-representation in the final sample is likely to result in a biased picture of wealth distribution. Moreover, wealth surveys generally include both complex and sensitive items. It follows that respondents are

[1] In 2015, the European System of Central Banks (ESCB) has established an expert group with the aim of comparing and harmonizing macro data (i.e. national accounts/financial accounts) and microdata (i.e. the Household Finance and Consumption Survey) on wealth.

not always able or even willing to report the correct amount of wealth they hold. Like non-response, measurement error is not at random and differs across population subgroups and portfolio items (D'Alessio and Neri, 2015; Ranalli and Neri, 2011).

The ideal solution for overcoming these problems would be to link survey data with administrative records (such as tax records or credit registers, as in Blanchet et al., 2018; Garbinti et al., 2018, 2020). Alternative approaches to data linkage are directly based on the use of wealth (tax) records (Alvaredo and Saez, 2009; Atkinson, 2016) or capital income information from tax records to construct wealth estimates assuming certain rates of return on assets (Saez and Zucman, 2016). Unfortunately, when such administrative records exist and are not limited in scope, they are not usually available for confidentiality reasons.

Because of that, the recent literature has developed methods to re-estimate the wealth distribution after combining survey data with the limited external information publicly available, such as aggregate figures from national accounts or lists of rich individuals' total wealth. This is usually achieved by fitting a Pareto distribution to the combined data. Vermeulen (2018, 2016) uses Forbes annual World's Billionaires List in combination with wealth surveys to estimate the total wealth held by rich households. He shows that the use of such lists increases the quality of the results (compared to estimating a Pareto model from survey data alone). Similarly, Jenkins (2016) estimates income inequality in the UK by fitting a Pareto distribution to a combination of tax data and survey data.

Building on these approaches, Chakraborty et al. (2019), Waltl (2021), and Chakraborty and Waltl (2018) extend the analysis by benchmarking survey results to the national accounts. It is important to highlight that the method used in these studies is not an imputation method, as rich lists are only used to aid in the estimation process. However, in a recent study, Bach et al. (2019) also implement these methodologies to impute rich list data to wealth surveys.

The common assumption behind all these studies is that unit non-response of wealthy households is the only reason for the micro-macro gap. This paper adds to the literature that tries to produce distributional indicators of wealth consistent with the national accounts, by proposing a methodology that draws on existing and well-established methods. We contribute to this literature in four ways.

First, whereas previous studies focus only on the missing part of the tail, assuming existing survey observations as representative, we claim that differential non-response also affects the representativeness of existing survey observations, which in turn affects estimates for the total number of households in the Pareto tail, and their total wealth. We propose a correction for differential non-response that accounts for the missing rich, but focuses on observed survey households. This correction does not substitute the imputation (Bach et al., 2019) or simulation (Waltl, 2021) procedures developed in the literature, but rather complements them by allowing for the correction of non-response bias among existing survey observations.

Second, while existing papers only focus on non-response at the tail of the distribution, we present a methodology that allows us to correct for measurement error as well. Dealing with both aspects simultaneously is important, even when the research purpose is to estimate the share of total wealth held by wealthy households. Indeed, some rich households may misreport their true wealth and therefore could be misclassified in the adjustment process. One advantage of our approach is that it enables us to compute distributional indicators that refer to "non-rich" households, such as those relating to financial vulnerability.

Our third contribution is that even if we apply well-established methods (such as Pareto modelling, imputation, and calibration) we show how to combine and use them in a single framework and how to evaluate the precision of the results. In particular, we propose the use of calibration as an imputation method, that generalizes and encompasses proportional allocation as a particular case.

The fourth and final contribution is to produce a modified and readily usable dataset in which survey values are adjusted for the above-mentioned quality issues and, by construction, the totals add up to the national accounts. While the existing papers are mainly focused on methods to estimate total wealth held at the top of the distribution, our adjusted dataset can be used for estimating any distributional indicator that may be of interest.

The paper is structured as follows. Section 2 describes the data sources used in and motivating our application. Section 3 presents the Pareto approach (Subsection 3.1), calibration (Subsection 3.2), and the methodology we propose to combine the two (Subsections 3.3, 3.4, 3.5). Section 4 describes the tools used to assess the properties of the proposed methods. Section 5 describes how the method applies to our data, while Section 6 discusses the results and the main findings of the application. Finally, Section 7 concludes.

2 Data

This paper uses the Household Finance and Consumption Survey (HFCS) and two sources of auxiliary information, namely, the national accounts, which include both financial and non-financial accounts, and rich list data.

The HFCS is a joint project of all the national central banks (NCBs) of the Eurosystem and several national statistical institutes (NSIs). The survey collects detailed household-level data on various aspects of household balance sheets and related economic and demographic variables, including income, private pensions, employment, and measures of consumption. The HFCS is conducted using a decentralized approach. A group of experts from the European Central Bank (ECB) and the NCBs (the Household Finance and Consumption Network, HFCN) coordinates the project, ensuring the cross-country comparability of the final data.

We use the second wave of HFCS (2014) and we restrict our analysis to four countries: Italy, France, Germany, and Finland. This choice is motivated by two considerations. First, rich lists and non-financial accounts are available for this subset of countries. Second, these surveys present methodological differences that can be used to evaluate our method. For example, some countries over-sample rich households using individual tax records (as in the French and Finnish survey) or using the information at the regional level (as in the German one), while others do not over-sample (as in the Italian case). Moreover, in some instances, the survey is linked to administrative data (as in the Finnish one).

In both cases of over-sampling and use of administrative records, we should expect the adjustment method to have a lower effect.

Our variable of interest is household net wealth defined as the sum of deposits, bonds, shares, mutual funds, money owed to the household, the value of insurance policies and pension funds, business wealth, and housing wealth, minus debts.

The second source of information is provided by national accounts. The financial component (financial accounts) is produced by NCBs and reports the total financial assets and liabilities held by households, classified by financial instrument and in order of liquidity based on the original maturity and negotiability (cash, deposits, insurance, and pension instruments). Non-financial accounts are produced by NSIs and contain the total value of dwellings, other buildings and structures, and land owned by households. Even if national account figures may suffer from quality issues and may adopt different concepts and definitions from those used in the survey, we employ them as a benchmark to correct survey data.

Rich lists are our third source of information. They have already been used in the literature to adjust for missing rich households (Vermeulen, 2018; Chakraborty and Waltl, 2018). Their use may generate concerns since the methodology adopted to populate them is often obscure and usually only figures for net worth are provided, with no financial instrument breakdown. Some studies have tried to resolve these issues by using different types of Pareto adjustments (Blanchet et al., 2017; Waltl, 2021). Others (such as Schröder et al., 2019) have also explored new ways of sampling high-wealth individuals with adequate precision. However, these methods can only be employed in specific instances when information on these households exists and is easily accessible. When these sources are not available, rich lists remain a reliable alternative, and evidence from Waltl (2021) indicates that, after rich lists have been integrated in the estimation of the Pareto tail, there might be little difference between the wealth estimated by different Pareto adjustments.

In our case, we use wealthy household data from "Forbes 2014 World's Billionaires List". This information has been replaced by that from larger region-specific lists, such as 2014 editions of Challenges' "Les 500 plus grandes fortunes de France" for France, Manager Magazin's list for Germany and Arvopaperi's list for Finland, when available. We also adjust this rich list data by estimating the debts and portfolio composition, based on portfolio shares from top wealth observations in the HFCS.[2] In this way, estimates for portfolio compositions among top fortunes can be obtained, and rich list data can be fully integrated with the HFCS for estimation purposes.

3 Methodology

Let w be household net wealth, w_i the net wealth for each individual household i in the population, for $i = 1, \ldots, N$, and $t(w) = \sum_{i=1}^{N} w_i$ be the population total to be estimated using survey data. Let $\mathcal{S}$ be the set of units selected in the sample and $\mathcal{S}_0$ be the final set of respondents. Let $\hat{t}(w) = \sum_{i \in \mathcal{S}_0} d_i w_i$ be the Horvitz-Thompson estimator, where d_i is

[2] This is a simplifying assumption. An improvement over this form of portfolio allocation is offered by the approach used in Chakraborty and Waltl (2018).

the sampling weight adjusted to account for smaller sample size of respondents, i.e. such that $\sum_{i \in \mathcal{S}_0} d_i = \hat{N}$.

Because of unit non-response and measurement error, the expected value of the Horvitz-Thompson estimator $\hat{t}(w)$ is generally lower than $t(w)$. Unit non-response occurs when some households refuse to participate to the survey. If this decision is related to household wealth (i.e wealthier households are more difficult to enlist in the survey than others), then the sample of respondents $\mathcal{S}_0$ may not adequately represent the upper tail of the distribution. Measurement error happens when the information collected in the survey w_i is different from the true unknown value w_i^*, for $i \in \mathcal{S}_0$. The error term $(w_i^* - w_i)$ may depend on many factors such as the difficulty of respondents to recall the required information or their unwillingness to report their true wealth.

Our methodology to address these issues is based on two techniques that are well-established in the literature. We use the Pareto distribution to compensate for unit non-response of wealthy households (Subsection 3.1), and the calibration approach, that is commonly used in survey sampling to deal with design weight adjustments, to address here measurement error (Subsection 3.2).

The two correction methods are dependent on each other and they must be implemented simultaneously. The Pareto correction starts with an assessment of the rich households available in the survey. Because of measurement error, some households could be misclassified and therefore a preliminary calibration adjustment is required. On the other hand, calibration is used again as an imputation method for the adjustment of measurement error across the whole distribution, requiring that the survey adequately represents the upper tail of the distribution.

In order to conduct the two adjustments simultaneously, we propose to run them in an iterative process, based on the procedure described in the following sections. The final product of the methodology is an adjusted survey data set with total estimates of net wealth, real assets, financial assets, and liabilities that match the aggregate figures in the national accounts balance sheet. This data set can be used to compute several distributional indicators of interest.

Before applying the method, we reclassify some definitions of wealth items used in the survey data in order to remove as many of the conceptual differences with national accounts as possible (see for instance EG-LMM, 2017; Chakraborty et al., 2019). In particular, we remove the wealth held by non-profit institutions serving households (NPISHs) from national accounts totals, and we only focus on the items with the highest level of comparability.

3.1 Pareto tail estimation

Wealth is Pareto distributed if, above a certain wealth threshold w_0, the complementary cumulative distribution (CCDF) of wealth is approximated by a power law, which, for $w_i \geq w_0$, can be expressed as:

$$P(W \leq w_i) = 1 - (w_0/w_i)^{\alpha}, \quad (1)$$

where the parameter $\alpha \in \mathbb{R}^+$ indicates the shape of the tail. The lower the value of α, the fatter the tail, and the more concentrated the wealth is.

If survey data were perfectly representative of a population, α could easily be estimated via maximum likelihood and w_0 could also be identified using the properties of the mean excess function from Yang (1978). However, in the presence of differential non-response, some households have zero probability of being included in the survey (the missing tail, from now on) after wealth reaches the upper wealth bound in the survey w_t (the truncation point, from now on). This may be due to the difficulties in contacting such rich households to even negotiate an interview, or to a specific decision by the data producer to exclude them for operative or confidentiality reasons. Also, it is important to recall that this truncation in not "sharp" in a strict sense: the presence of differential non-response implies that observed households in the Pareto tail are also under-represented as the probability of a household being interviewed approaches zero the closer its wealth is to the truncation point, and that this smoothly decreasing response rate will be reflected in a smoothly decaying weighted density of the observed distribution in the Pareto tail (Vermeulen, 2018). The main implication of the presence of differential non-response is that conventional methods for the estimation of the α parameter will not work.

The method described in Vermeulen (2016, 2018) has been developed precisely to overcome these issues. The first step of the adjustment consists in selecting the threshold w_0. Previous research has often adopted the arbitrary threshold of €1 million. We also adopt this threshold, but also provide robustness estimates with a lower 500k threshold for Italy, justified by the significantly higher level of under-reporting and differential non-reponse.

Accordingly, define $\mathcal{S}_T = \{i \in \mathcal{S}_0, \text{ s.t. } w_i \geq w_0\}$ as the sub-sample of m_T respondents whose wealth is above the threshold. Let $\mathcal{S}_R$ be the set of population units in the rich list. We assume that for all units in $\mathcal{S}_R$, w_i is larger than the maximum value observed in $\mathcal{S}_T$.[3] Then, $\mathcal{S}_R$ and $\mathcal{S}_T$ are appended creating a new set $\mathcal{S}_I = \mathcal{S}_T \cup \mathcal{S}_R$ of dimension m_I. In our approach, the imputed set $\mathcal{S}_I$ is only used for the estimation of α, and afterwards all adjustments are only applied to the original sample of respondents $\mathcal{S}_0$.

Let $\bar{D}_I$ be the average survey weight of all units in $\mathcal{S}_I$, that is $\bar{D}_I = \sum_{i \in \mathcal{S}_I} d_i/m_I$, where we consider $d_i = 1$ for $i \in \mathcal{S}_R$. Now, let $\mathcal{S}_j$ be the subset of $\mathcal{S}_I$ with the j-th wealthiest households, for $j = 1, \ldots, m_I$. That is $\mathcal{S}_1$ is made of the wealthiest household, $\mathcal{S}_2$ is made of the two wealthiest households, and so forth, so that $\mathcal{S}_{m_I} = \mathcal{S}_I$. Then, $\bar{D}_j$ is the average weight of the j wealthiest households, i.e. $\bar{D}_j = \sum_{i \in S_j} d_i/j$. Linear estimates for α can then be obtained through the following least squares specification (see also Gabaix and Ibragimov, 2011):

$$\ln[(j - 1/2)\bar{D}_j/\bar{D}_I] = C - \alpha \ln(w_j), \tag{2}$$

where $C = \ln(m) + \alpha \ln(w_0)$ and w_j is the net wealth of the j-th wealthiest household.

Using Monte Carlo simulation, Vermeulen (2018) has shown that this method can produce estimates that approximate the true population parameter, while accounting for the

[3] We then remove from the rich lists all observations displaying lower wealth than what the maximum wealth observed in the survey. This is an extremely rare occurrence, and only few observations from the Finnish rich list were removed..

increase in non-response probability as wealth increases. This means that information on the Pareto parameter can be retrieved even when survey weights decay progressively until the probability of non-response reaches 1. The method has been replicated in Chakraborty and Waltl (2018) and Waltl (2021), who showed that this estimator can produce unbiased and consistent estimates of α even when information on top tail observations is obtained from commonly available rich lists.

The second step of the adjustment consists in estimating the total wealth in the top tail $t(w; top)$ by multiplying the estimate of the total number of rich households from the survey, $\hat{N}_T = \sum_{i \in \mathcal{S}_T} d_i$, by the mean of the estimated Pareto distribution, given by $\alpha w_0/(\alpha - 1)$, for $\alpha > 1$. The adjustment proposed by Vermeulen (2016; 2018) stops here: after this estimate is produced, no further adjustment is made to the survey. Our intention, however, is to use this information to adjust survey data. For our intents and purposes, this estimate could be used to calibrate the sampling weights of rich households in the survey to the total wealth implied by the Pareto adjustment. However, this approach assumes, for simplicity, that the sample estimate of N_T (the total number of households whose wealth exceeds w_0) is unbiased. We believe that this assumption should be relaxed. After all, as discussed earlier, the presumption of differential non-response is based on the assumption that a number of very rich households could not be sampled. The absence of these households leads not only to the missing tail, but also to a smooth under-estimation of survey weights as wealth increases. As these households are all located in the higher part of the Pareto tail, even the undersampling of a small number of households could have a sizeable impact on the amount of estimated wealth in the tail.

As a result of the underestimation of households in the Pareto tail, estimates for total wealth in the tail will also be underestimated. If we are able to re-estimate the Pareto tail parameter, we should also be able to reassess the total number of households in the Pareto tail. We then propose a novel method for the estimation of the number of missing rich households and their wealth.

Recall that w_t is the truncation point above which there are no rich households in the sample. Let's focus here on $\mathcal{S}_T$ only and let $\mathcal{S}_j$ be now the subset of $\mathcal{S}_T$ with the j-th wealthiest households, for $j = 1, \ldots, m_T$. That is $\mathcal{S}_1$ is made of the wealthiest household in $\mathcal{S}_T$, $w_{m_T} = w_t$, and $\mathcal{S}_{m_T} = \mathcal{S}_T$. Following from the Glivenko-Cantelli theorem, because of truncation, the empirical cumulative distribution function resulting from this sample is unlike the theoretical distribution implied by the Pareto adjustment. In particular, the following relation does not hold:

$$\frac{\hat{N}_T - \hat{N}_{j-1}}{\hat{N}_T} - \left[1 - \left(\frac{w_0}{w_j}\right)^{\alpha}\right] \approx 0 \tag{3}$$

where $\hat{N}_{j-1} = \sum_{i \in S_{j-1}} d_i$, that is the sum of weights for all households with wealth greater than w_j, so that $\hat{N}_T - \hat{N}_{j-1}$ is an estimate of number of households in the population whose wealth is between w_0 and w_j. This relation means that the empirical CDF will always suffer from a bias equal to or larger than zero since units whose wealth exceeds w_t are unobserved. Substituting $\hat{N}_T$ with the unknown N_T in equation (3) and re-arranging leads to the following equation

$$N_T \approx \frac{\hat{N}_T - \hat{N}_{j-1}}{1 - (w_0/w_j)^\alpha}. \tag{4}$$

Analytically, the estimate from equation (4) should be the same for each $j \in \mathcal{S}_T$. In practice, with empirical data, variability in survey weights will affect the estimate of the number of households in the tail. Because of differential non-response, this becomes a particularly thorny problem when weight quality can deteriorate as the observed wealth gets closer to the truncation point w_t, where the maximum wealth in the survey sample $\mathcal{S}_T$ is recorded. The results can then be improved by estimating N_T for each value of wealth of top tail observations and then getting the average. In particular,

$$\tilde{N}_T = \frac{1}{m_T} \sum_{j=1}^{m_T} \frac{\hat{N}_T - \hat{N}_{j-1}}{1 - (w_0/w_j)^\alpha}. \tag{5}$$

An estimator of the number of missing, unobserved, households after the truncation point can be computed as $\hat{N}_t = \tilde{N}_T(w_0/w_t)^\alpha$. To account for these missing households, the total number of observable households can be estimated as

$$\hat{N}_{obs} = \tilde{N}_T[1 - (w_0/w_t)^\alpha]. \tag{6}$$

Finally, the total wealth in the top tail $\hat{t}(w; top)$ can be estimated by the product of the estimated number of households and the Pareto mean as

$$\hat{t}(w; top) = \tilde{N}_T \frac{\alpha w_0}{(\alpha - 1)}. \tag{7}$$

Wealth in the missing part of the tail can similarly be computed as $\hat{N}_t \alpha w_t/(\alpha - 1)$, setting the new threshold at the truncation point w_t. Note that this is possible because the Pareto shape parameter does not change along the Pareto distribution.

3.2 Calibration

Calibration is a method whose aim is to correct the sampling weights d_i through re-weighting methods while keeping the individual responses w_i unchanged (Deville and Särndal, 1992; Särndal, 2007). In the literature, this approach is mainly used: (i) to force consistency of certain survey estimates with known population quantities; (ii) to reduce non-sampling errors such as non-response errors and coverage errors; (iii) to improve the precision of estimates (Haziza et al., 2017).

Calibration is achieved through the following optimization problem for finding a new set of weights $d_i^* = d_i a_i$:

$$\min_{d_i^*} \sum_{i \in \mathcal{S}_0} G(d_i^*; d_i) \quad s.t. \quad t(z) = \sum_{i \in \mathcal{S}_0} d_i^* z_i, \tag{8}$$

where $G(d_i^*; d_i)$ is a distance function between the basic design weights and the new calibrated weights, z_i is the value on unit i taken by a (possibly) vector valued auxiliary variable z, and $t(z)$ are the benchmark constraints, that is the known vector of population totals or counts of the calibration variables z. The adjustment factors a_i's are a function of the z_i's and they are computed so that final weights meet benchmark constraints, $t(z)$, while, at the same time, being kept as close as possible to the initial ones. Closeness can be defined by means of several distance functions (see Table 1 in Deville and Särndal, 1992), the most common being the chi-squared type,

$$G(d_i^*; d_i) = \frac{(d_i^* - d_i)^2}{d_i c_i} \tag{9}$$

where c_i are known constants whose role will be discussed in more detail later, for which an analytical solution always exists.

The final output is a single new set of weights to be used for all variables. The magnitude of the adjustment factors and therefore the variability of the final set of weights is a function of the number of constraints, i.e. the length of $t(z)$, and the imbalance (the difference between the Horvitz-Thompson estimate and the population total of z). Very variable weights hinder the quality of final estimates for sub-populations and for variables that are not involved in the calibration procedure. For these reasons, weights are usually required to meet range restrictions such as to be positive and/or within a chosen range. This can be achieved by carefully choosing and tuning the distance function $G(\cdot)$.

The method was originally proposed to improve the efficiency of the estimators and to ensure coherence with population information, but then it was also largely applied to adjust for non-response (Särndal and Lundström, 2005). For example, Little and Vartivarian (2005) showed that if the variables used to construct the weights are associated both with non-participation and with the variable of interest, the bias and the variance of the estimator are reduced.

The main problem with the use of household balance sheet data in re-weighting methods is that wealth is generally skewed and concentrated in the hands of a small group of the population that has both low propensity to participate in the survey and different socio-demographic characteristics from the average population.

3.3 Adjusting for non-response: Pareto-calibration

We begin by exploiting the information obtained after fitting a Pareto distribution, as in Subsection 3.1, to adjust the wealth distribution in the survey for differential non-response using the calibration methods described in Subsection 3.2. We use w_0, $\hat{\alpha}$ and Equation (5) to estimate the total number of observable households over the threshold $\hat{N}_{obs}$ (see Equation (6)) and their total net wealth by

$$\hat{t}(w; obs) = \hat{t}(w; top) - \hat{t}(w; miss),$$

where $\hat{t}(w; miss) = \tilde{N}_T \alpha w_t / (\alpha - 1)$. We then obtain the corresponding figures $\hat{t}(w; bot)$ and $\hat{N}_{bot}$ for households below the threshold w_0. We then calibrate the sampling weights from sample $\mathcal{S}_0$ using the following constraints,

$$t(z) = (\hat{t}(w; obs), \hat{N}_{obs}, \hat{N}_{bot}, \hat{t}(y; bot), t(x)) \tag{10}$$

where $\hat{t}(y; bot)$ is a vector of Horvitz-Thompson estimators decomposing the initial wealth of observations below the threshold into their corresponding portfolio items, and $t(x)$ is a vector of population counts for demographic characteristics.

After calibrating survey data to these parameters, we obtain non-response adjusted weights d_i^*'s. This approach will be referred to as 'Pareto-calibration' from now on. It is worth noting that our Pareto-Calibration approach is also intended to ensure that the transition between the Pareto tail and the rest of the distribution remains smooth. As will be discussed later, this property is confirmed from figures 2 to 5. Given the calibration constraints, wealthier households will see the largest increase in their weights, whereas the weights of households near the threshold will remain mostly unchanged. This will be shown later in Figure 1.

Should the survey be suffering from differential non-response issues only, this step might be sufficient to fill the gap with the financial accounts. However, this is not always the case: provided that we have a good approximation of wealth distribution in the tail, the remaining differences in coverage between the estimate obtained in Equation (7) and the national accounts will then be left to measurement error.

3.4 Adjusting for non-response and measurement error: Simultaneous approach

In order to correct for measurement error, we combine the adjustment for differential non-response described in Subsection 3.3 with the following procedure. First, we run the Pareto-calibration adjustment, as described earlier. Let d_i^*, for $i \in \mathcal{S}_0$, be the final weight from the non-response adjustment procedure. Next we run a calibration procedure as in Equation (8) in which (i) the d_i^*'s are now the starting weights and (ii) the set of benchmark constraints $t(z)$ are given by the vector of macro aggregates of wealth items y. The adjustment factor a_i, for $i \in \mathcal{S}_0$, obtained by this procedure is such that

$$\sum_{i \in \mathcal{S}_0} d_i^* a_i y_i = t(y) \tag{11}$$

We apply this adjustment factor directly to the variables of interest so that

$$y_i^* = a_i y_i, \tag{12}$$

that is, we choose to adjust the values of observations for the components of wealth, rather than their weights. In this way, we avoid the possibility of large adjustments to affect weights that are used to compute estimates for all variables in the survey, and not only for wealth components. This approach shares similar traits with reverse calibration introduced by Chambers and Ren (2004) to deal with outlier-robust imputation.

Is it worth recalling that y_i is vector-valued. Note, also, that this calibration is multivariate because it accounts for all constraints with respect to macro estimates in a single procedure and, therefore, it accounts for the multivariate structure of the variables included in y. In addition, every household has a different adjustment factor a_i that depends on all the values of y. In this way, rather than rescaling individual components of wealth by a different factor (which would be the same for all households), we use the same scaling factor for all components, but a different factor for each household.

Proportional allocation, which consists of allocating the gap by multiplying each component of y_i by the corresponding inverse of the item-specific coverage ratio, can be embedded in our approach as it can be seen as a particular case of univariate calibration. In fact, if we focus on a single item, y_1, the adjustment factor used by proportional allocation can be obtained as the solution to a univariate calibration procedure in which (i) the starting weights are again the d_i^*'s, (ii) there is only one benchmark constraint $\sum_{i \in \mathcal{S}_0} d_i^* a_i y_{i1} = t(y_1)$, and (iii) the distance function $G(\cdot)$ is chi-squared as in (9) with constants $c_i = 1/y_{i1}$. The proof is omitted for brevity, but it is close in spirit to Example 1 in Deville and Särndal (1992).

This equivalence sheds some light on the role of the constants c_i's in the distance function (8). In univariate calibration, if they are chosen to be the inverse of the variable in the constraint, then the adjustment factors are shrunk towards a common value for all households as in proportional allocation. On the contrary, if they are set to be constant, the adjustment factors would be roughly proportional to the values of the item. For this reason, in the proposed multivariate calibration for imputation, we have set the constants to possibly depend on the wealth of the household, that is

$$c_i = \left(\frac{1}{w_i}\right)^{\tau}, \tag{13}$$

where $\tau \geq 0$ can be seen as a shrinkage factor: larger values provide adjustment factors that are more uniform across households, while values towards 0 provide adjustment factors with a higher variability and correlation with w_i.[4]

Note that the adjustment factors a_i may be very variable because we are using a multivariate calibration approach and imbalance due to measurement error can be considerable. This is particularly relevant when using the chi-squared distance function for which the adjustment factors can take even negative values. For this reason, we recommend the use of alternative distance functions, such as the raking (Case 2 in Deville and Särndal, 1992) for which positive adjustments are ensured, or the range restricted version of the chi-squared and of the raking distance functions (Cases 6 and 7 in Deville and Särndal, 1992), for which adjustments are bounded to be in a pre-specified interval.

In order to account for the missing wealthy households, we add a single observation with weight $\hat{N}_t$ and wealth $\hat{t}(w; miss)/\hat{N}_t$. This observation's portfolio is also allocated using portfolio shares in the Pareto tail of the distribution.

At the end of the multivariate calibration the gap is filled. However, the distribution of w_i has changed, because its components have changed. Some households, which were initially classified as not rich (below the threshold) may have moved into the top tail of wealth distribution. Therefore, we need to find the new Pareto threshold, and we do so by rescaling the threshold w_0 by the average adjustment factor. This requires an iterative procedure which alternates a Pareto-calibration step that improves coverage with a multivariate calibration step that addresses measurement error. The two steps are iterated

[4]For this work, we set $\tau = 1$. Future research might seek to retrieve information on τ using external data where no misreporting behaviour is present.

until convergence. Convergence has been set on the parameter α of the Pareto distribution: if the estimated values in two consecutive steps differ by less than a small predefined threshold the procedure stops [5].

3.5 A special case: Single-iteration approach

If one is willing to assume that (i) the relative measurement error is independent from the observed wealth, at least among the very rich, converging in probability to a constant ζ, i.e. $(w^* - w)/w \xrightarrow{p} \zeta$, so that, on average, the unobserved 'true' total wealth will be given by $\hat{w}_i^* = \zeta w_i$, then the method simplifies. Another important assumption is that (ii) the degree of relative measurement error between survey and rich list is also similar. Should these assumptions hold, survey wealth would still be Pareto distributed with tail parameter α after adjusting for measurement error. It follows that total wealth in the survey would scale up to $\sum_{i \in \mathcal{S}_0} \zeta d_i^* w_i$, and the Pareto CDF would turn into $F_\alpha(\zeta w_i) = 1-(\zeta w_0/\zeta w_i)^\alpha$. Simplifying this last formula and updating equation (7) for measurement error, we obtain the following estimate for total wealth:

$$\zeta \hat{t}(w) = \zeta \left(\frac{\alpha w_0}{(\alpha - 1)} \hat{N}_{obs} + \sum_{i \in \mathcal{S}_0} d_i^* w_i \right). \tag{14}$$

This implies that our estimate for α does not depend on the scaling of the variables. In this case, the coefficient for the Pareto-adjusted coverage ratio, given the national accounts total wealth, as in $\zeta = t(w)/\hat{t}(w)$, will yield the scalar to which to re-allocate reported survey wealth. To account for the missing wealth, wealth should be scaled to $\zeta(\hat{t}(w) - \hat{t}(w, miss))$, which, after Pareto-calibration, simplifies to $\zeta \sum_{i \in \mathcal{S}_0} d_i^* w_i$.

If the Pareto shape parameter is unaffected by the re-scaling, the iterative procedure is no longer needed. The adjustment for measurement error and for non-response at the tail of the distribution can be run independently of each other.

Ideally, if the above mentioned assumptions hold, whatever the adjustment method for measurement error is used, the final data should still be Pareto distributed among rich households. In practice, we adopt a single-iteration approach by employing either proportional allocation or the non-iterative version of the multivariate calibration method described in 3.4.

As will be discussed later, the robustness of these estimates strongly depend on the level of discrepancy in terms of measurement error between survey and rich list data. The assumption (ii), in fact, does not hold in all contexts.

4 Assessment of the method

The ideal approach for assessing the quality of the results would be to compare them with an external benchmark, for instance, from highly reliable administrative records. Without such auxiliary information, we can assess the method in two ways. First, we

[5] It is worth stressing that the convergence of the process could also not be achieved, especially if the gap to be filled is sizeable.

assess the robustness of our results by comparing them with other estimators based on different assumptions. Second, we assess the precision of our results by estimating their variability.

Beyond our simultaneous approach, we compute four alternative estimators:

- 'Survey & missing tail'. The results are produced using the unadjusted survey data, plus an estimation of the total wealth held by rich household with zero probability of being in the survey (missing tail).
- 'Pareto-calibration & missing tail'. Survey data are adjusted with the Pareto-calibration model. Survey weights are calibrated using constraints from Equation (10) and the total wealth of the missing tail is included in the estimate.
- 'Pareto-calibration, proportional allocation & missing tail'. This method adds to the previous one a correction for measurement error based on proportional allocation, as in Fesseau and Mattonetti (2013). This is a naive method based on the assumption that measurement error is equal across households and that it only depends on the instrument.
- 'Single-iteration approach & missing tail'. In this method, the correction for measurement error is only iterated once, as described in Subsection 3.5

Variance estimation in our methodology has two main components. The first one is the sampling variance, which indicates the variability introduced by choosing a sample instead of enumerating the whole population, assuming that the information collected in the survey is otherwise entirely correct. A second source of variability is imputation variance which refers to the fact that the methodology for filling the gap can produce several different plausible imputed data sets. The uncertainty due to the imputation process adds up to the sampling variance.

To estimate the overall variability we use the Rao-Wu rescaled bootstrap weights released with HFCS data to account for sampling variability (Eurosystem Household Finance and Consumption Network, 2020). For each of the 1,000 sets of bootstrap weights we replicate all the methods previously described. In each replication, the parameters of the Pareto distribution are re-estimated introducing additional variability. We then obtain the mean and standard deviation from all successful simulations[6] to evaluate the robustness of our methods and derive a measure of their variability.

5 Application to the HFCS

The method described in the previous sections has been applied to the second 2014 wave of the HFCS. The first step consists of estimating the α parameter of the Pareto distribution. The estimated coefficients are shown in Figure 1, which show how the method Vermeulen (2016) can significantly affect the magnitude of the estimated parameter. α indicates the Pareto shape parameter estimated by imputing the rich list, while θ shows

[6] A simulation is flagged as unsuccessful, and discarded, whenever a calibration procedure fails because of lack of convergence under the chosen restraints.

these estimation results with survey data only. The Figure also illustrates the outcome of the Pareto-calibration process, showing the empirical CCDF on a log-log scale before and after the adjustment. Re-weighted figures are produced by using the proposed Pareto-calibration method. Figures 2, 3, 4, and 5 also shows the empirical probability density distribution before and after the Pareto-Calibration step. Notably, no significant discontinuities between the Pareto tail and the rest of the distribution are to be noted, confirming that the re-weighting methodology can account for smoothly declining response rates without affecting the nature of the overall distribution. No noticeable discontinuities appear after the subsequent adjustments too.

Table 1 shows coverage ratios between survey wealth estimates and financial accounts. Column (1) shows initial coverage ratios, while Column (2) displays the coverage ratio for adjusted data, and Column (3) grosses up survey wealth by estimating total wealth after truncation and adding it to the previous estimate. Columns (4) and (5) show the estimated number of households in the Pareto tail, along with the number of "missing rich".

Overall, these figures suggest that the proposed Pareto-calibration approach can produce substantial improvements in survey coverage, especially in the absence of oversampling or administrative data. In the case of Finland and Germany, the discrepancies between micro and macro figures virtually disappear after calibrating survey weights and accounting for the unobservable households. These results suggest that corrections for differential non-response can solve the missing wealth problem alone in cases where measurement error is limited: the fact that the Finnish survey relies on registry data, where measurement error is virtually absent, and that the coverage gap is completely filled after calibrating the survey weights, provides evidence in support of this argument.

Also, it should be noted that the improvements in coverage between columns (1) and (2) are much larger than the improvements that could be noted between columns (2) and (3). This makes a strong case in favour of our Pareto-Calibration approach, meaning that simply addressing for the decay in survey weights in the observed part of the distribution can already address the largest part of the non-response problem.

Furthermore, having re-estimated the number of households in the Pareto tail of the survey, our method also shows substantial improvements in coverage over the grossing up methods already explored in the literature, and suggests that adjustments for non-response should also focus on correcting the number of households estimated to be in the Pareto tail, rather than only the wealth contained in it.

Coverage is also significantly improved for Italy and France, but the persistence of a mismatch between survey data and financial accounts points to the presence of measurement error. In the case of the Italian survey, a case could be for the 1 million empirical threshold being too high. The change in the Pareto α coefficient before and after the Pareto Calibration is significant, and suggests that the threshold is placed in a point where survey weights are already suffering from decay due to non-response. Considered that the Italian survey is already known for suffering from both non-response and measurement error, it is not unreasonable to believe that the threshold could be placed lower in the distribution. This is evident from a visual inspection of the wealth distribution, as can be seen from Figures 2 and 6.

As discussed, there is no agreed-upon method for the detection of the threshold, other than visual inspection. Some methods have been developed based on the Mean Excess Function from Yang (1978), but these methods fail to find a threshold when differential non-response is present. However, these methods can still provide a lower bound for the Pareto threshold. We find this lower bound to be EUR 310,084 for Italy, 567,378 for France, 254,000 for Germany and 880,806 for Finland. This suggests that the 1 million threshold is already a generous choice for the Italian survey, and provide robustness estimates for Italy with a EUR 500,000 threshold. Figures 6 and 7 show the distributional changes with this new threshold: these distributional changes now appear much more reasonable, with no significant change in the estimated Pareto parameter.

After dealing with the issue of nonresponse at the tail of the distribution, we use multivariate calibration to adjust for measurement error along the whole distribution.

As benchmark constraints $t(y)$ we use the financial instruments with high conceptual comparability between survey and financial accounts – namely, deposits, bonds, shares, funds, insurance products and liabilities – following from the comparability scale provided by EG-LMM (2017). The resulting adjustment factors are then applied to financial instruments with lower comparability – business and housing wealth – which, assuming that measurement error is comparable across similar financial instruments, should ensure that the adjustment will not be biased by the presence of instruments with low comparability.

We then iterate the Pareto-calibration and the multivariate calibration until there is convergence. Convergence has been set on the parameter α of the Pareto distribution: if the estimated value in two consecutive steps differs by less than a small predefined threshold,[7] the procedure stops. Convergence is usually achieved in a limited number of steps (between 1 and 3 in the application at hand).

Table 2 shows the average values of the adjustment factors a_i's (as well as coefficients of variation) as a function of gross wealth percentiles at the end of the iterative procedure for the four countries. That is, these are the overall adjustment of the survey variables at the end of the procedure obtained as the ratio between the final imputed values and the ones from the original survey.

6 Results

Table 3 shows the distributional results indicating the proportion of net wealth held by the top 1, 5, 10, and 20 percentiles, along with the bottom 50%. Gini inequality indices are also presented in Column (6), while Column (7) provides the estimated Pareto tail parameter α given the data. These figures are reproduced under each allocation method. The bootstrap-based standard deviation is reported in parentheses for each estimate.

The first set of rows ('Base Survey') presents distributional figures from the unadjusted HFCS data. As is well known, truncation in top wealth distribution and measurement error can cause survey estimates to understate the true level of wealth inequality, and the figures presented in the table provide support for this possibility. Indeed, estimates from the unadjusted HFCS would suggest that wealth inequality in Italy, which has

[7] In the current application, this tolerance was set at 0.05.

one of the largest micro-macro gaps, is close to the inequality level in Finland, where the gap is lower.

For all subsequent sets of rows, which correspond to the alternative estimators discussed in Section 4, we also include an adjustment for the unobserved part of the Pareto tail as presented in Subsection 3.1. To do this, these missing households are imputed as a single observation whose weight and wealth are respectively equal to the estimated number of unobserved households and the estimated average wealth in the unobserved Pareto tail.

The second set of rows ('Survey & missing tail') displays estimates produced using the un-adjusted survey data, plus the missing tail households. Such adjustment raises inequality levels for all the surveys considered. The magnitude of the increase depends on the size of the gap to be filled. For surveys, such as the Italian and German ones, in which truncation bias is particularly pronounced, the sole inclusion of these unobserved households increases the proportion of wealth held by the top 1% households by at least 10 percentage points, respectively. This increase is much less pronounced for the French and Finnish surveys, where the truncation is also much more modest.

Survey weights are then adjusted using the proposed Pareto-calibration method to produce the figures shown in the third set of rows ('Pareto-calibration & missing tail'). This method produces a further increase in inequality, even if the magnitude is smaller than the previous step.

The row sets from fourth to sixth adjust the survey by applying the estimators described in Section 4. For countries like Finland and Germany, whose gap with Financial Accounts is filled after employing the Pareto-Calibration method, measurement error seems to be a negligible issue and these adjustments might not be needed.

In the fourth set of rows ('Par-cal, proportional allocation & missing tail'), portfolio items are scaled proportionally to the Financial Accounts aggregates,after the Pareto adjustment.

The final two sets of rows show the results obtained when combining the Pareto method with the multivariate calibration approach, either in a single iteration (Par-cal, Single-iteration approach & missing tail) or through an iterative process (Simultaneous approach).

As expected, all the estimators point to the fact that inequality is underestimated in survey data. The increase in the estimated degree of inequality is proportional to the severity of both non-response among wealthy households and measurement error problems. For instance, the Gini index increases by 10 points in Italy when considering the Simultaneous approach. For the other countries, the increase ranges from 2 points in Germany to 5 points in France.

A second result is that the Pareto adjustment has a larger influence in determining the final inequality statistics compared to calibration. The estimators which add a further adjustment by calibration or proportional allocation produce very similar results to the Pareto-calibration & missing tail estimator. Table 3 also shows that the Pareto tail parameters estimated using a rich list are different from those based on the survey alone (column 7). Using the Pareto method without external information would therefore produce bias results.

Even if the missing wealthy households is the most important issue to tackle, we also find that also measurement error should be taken into account. This further adjustment generally reduces the inequality levels. For instance, the share of total wealth hold by the top 10 percent of the richest households decreases by about 5 points when considering the simultaneous approach instead of the Pareto-calibration estimator.

Another important result is that the simultaneous approach produces in some cases different results compared to the single-iteration approach. Therefore we don't find a strong support for the assumption that the relative error converges in probability to a constant, as discussed in section 3.5 .

This is the case for the Italian survey, where the Single Iteration approach leads to inflated and highly variable results. As discussed, we believe this also to be connected to the choice of a Pareto threshold. Results obtained under a lower threshold (provided in Table 4), are closer to the ones produces by the simultaneous approach (at least for some indicators such as the Gini index).

Finally, as to variance estimation, the adjustment methods generally produce a decrease in the reliability of the results as we reach the highest percentiles. This is expected as the sample size in increasingly smaller as we move upwards in the distribution of wealth. For each method, the precision increases when the statistic relates the bottom or median part of the wealth distribution. The estimators of the wealth share held by the top 1 percent have a lower precision in all countries.

Compared to other methods, the simultaneous approach produces in many instances the lowest increase in variability. This is also due to the use of multivariate calibration, a method that was originally developed to increase the precision of the estimators. The final coefficients of variation are not very different to those based on the unadjusted survey data, especially for the statistics that do not relate to the top tail of the distribution.

As a final exercise, it is worth comparing our results with other estimates from the literature based on other methods.

Our results show a higher level of wealth inequality compared to those obtained by Vermeulen (2018).

are broadly comparable to the estimates for financial assets shown in Vermeulen (2016).

Looking at studies using different approaches, Garbinti et al. (2020) and Acciari and Morelli (2020) have recently reconstructed the wealth distribution of France and Italy, respectively. In both cases, some differences between the estimates of the wealth held by the top 1% of the distribution are to be noted. These differences are in the order of around 5-6 percentage points. This could be explained either by a difference in the definition of wealth, by the higher variability of top 1% estimates, by the relatively poor quality of rich lists in general (which might overestimate wealth when compared to administrative or survey data), or a combination of these factors.

Nonetheless, differences between these estimates and ours seem to disappear once we look at other distributional indicators: in both cases, estimates for the wealth held by the top 10% of the distribution are remarkably similar, and similar results are to be noted across the rest of the distribution.

7 Conclusions

In this paper, we show how a combination of well-established methodologies for the fitting of a Pareto distribution and the calibration of survey data can be used to correct for non-response and misreporting when only limited external information is available.

We apply these methods to the HFCS data, using the 2014 Finnish, French, German, and Italian surveys, and employing rich list data from Forbes or national press sources, along with household sector aggregates from national accounts, as auxiliary sources of information.

We find that, in line with previous studies, inequality estimates from the survey data understate the population parameters, depending on the severity of both non-response and measurement error. We also find that for some countries the adjustment for the missing rich households accounts for most of the gap between micro and macro figures. Moreover, we present a method to assess the precision of our results. As expected, all the adjustment methods increase the variability of the estimators. Nonetheless, the method we propose produces the lowest increase.

Finally, as a by-product of our methodology, we obtain an adjusted micro dataset which allows us to compute many distributional indicators, besides the ones shown in this paper.

Additional work is still needed for the refinement of the methodology we propose. For example, the estimation of the number of wealthy households could be further validated and improved, for instance by using alternatives to rich lists (such as tax records) or by applying additional methods (such as the Type II Pareto or the Estate Multiplier Method). Also, the correction of measurement error could be further improved by enriching the auxiliary granular vector with more granular external information (such as administrative records).

It is worth stressing that our method can be easily adapted in case additional external information is available. Indeed, both the Pareto-calibration adjustment and the multi-variate calibration methods can be enhanced with external information and can be run separately when needed.

Acknowledgements

The paper has greatly benefited from our discussions with all members of the EG-LMM and EG-DFA working groups.

References

Acciari, P. and Morelli, S. (2020). *Wealth Transfers and Net Wealth at Death: Evidence from the Italian Inheritance Tax Records 1995–2016*. University of Chicago Press.

Alvaredo, F., Atkinson, A. B., and Morelli, S. (2018). Top wealth shares in the UK over more than a century. *Journal of Public Economics*, 162:26–47.

Alvaredo, F. and Saez, E. (2009). Income and wealth concentration in spain from a historical and fiscal perspective. *Journal of the European Economic Association*, 7(5):1140–1167.

Ampudia, M., van Vlokhoven, H., and Żochowski, D. (2016). Financial fragility of euro area households. *Journal of Financial Stability*, 27:250 – 262.

Arrondel, L., Lamarche, P., and Savignac, F. (2019). Does inequality matter for the consumption-wealth channel? empirical evidence. *European Economic Review*, 111:139–165.

Atkinson, A. B. (2016). Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Economica*, 84(334):129–156.

Bach, S., Thiemann, A., and Zucco, A. (2019). Looking for the missing rich: tracing the top tail of the wealth distribution. *International Tax and Public Finance*, 26(6):1234–1258.

Blanchet, T., Flores, I., and Morgan, M. (2018). The weight of the rich: Improving surveys using tax data. *WID.world WORKING PAPER SERIES N° 2018/12*.

Blanchet, T., Fournier, J., and Piketty, T. (2017). Generalized pareto curves: theory and applications. *WID.world WORKING PAPER SERIES N° 2017/3*.

Casiraghi, M., Gaiotti, E., Rodano, L., and Secchi, A. (2018). A "reverse Robin Hood"? The distributional implications of non-standard monetary policy for Italian households. *Journal of International Money and Finance*, 85(C):215–235.

Chakraborty, R., Kavonius, I., Perez-Duarte, S., and Vermeulen, P. (2019). Is the top tail of the wealth distribution the missing link between the household finance and consumption survey and national accounts? *Journal of official Statistics*, 35:31–65.

Chakraborty, R. and Waltl, S. R. (2018). Missing the wealthy in the HFCS: micro problems with macro implications. ECB Working Paper Series No 2163, European Central Bank.

Chambers, R. L. and Ren, R. (2004). Outlier robust imputation of survey data. *The Proceedings of the American Statistical Association*, pages 3336–3344.

Christelis, D., Georgarakos, D., Jappelli, T., Pistaferri, L., and van Rooij, M. (2021). Heterogeneous wealth effects. *European Economic Review*, 137:103805.

Coibion, O., Gorodnichenko, Y., Kueng, L., and Silvia, J. (2017). Innocent bystanders? monetary policy and inequality. *Journal of Monetary Economics*, 88:70 – 89.

Colciago, A., Samarina, A., and de Haan, J. (2019). Central bank policies and income and wealth inequality: A survey. *Journal of Economic Surveys*, 0(0).

D'Alessio, G. and Neri, A. (2015). Income and wealth sample estimates consistent with macro aggregates: some experiments. Questioni di Economia e Finanza (Occasional Papers) 272, Bank of Italy, Economic Research and International Relations Area.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418):376–382.

EG-LMM (2017). Understanding, quantifying and explaining the differences between macro and micro data of household wealth: Final report. mimeo, European Central Bank.

Eurosystem Household Finance and Consumption Network (2009). Survey data on household finance and consumption research: summary and policy use. ECB Occasional Paper Series No 100, European Central Bank.

Eurosystem Household Finance and Consumption Network (2020). The household finance and consumption survey:methodological report for the 2017 wave. ECB Statistics Paper Series No 35, European Central Bank.

Expert Group on Linking Macro and Micro data (2020). Understanding household wealth: linking macro and micro data to produce distributional financial accounts. Statistics Paper Series 37, European Central Bank.

Fesseau, M. and Mattonetti, M. L. (2013). Distributional measures across household groups in a national accounts framework. OECD Statistics Working Papers No. 2013/08, Organisation for Economic Co-Operation and Development (OECD).

Frémeaux, N. and Leturcq, M. (2020). Inequalities and the individualization of wealth. *Journal of Public Economics*, 184:104145.

Gabaix, X. and Ibragimov, R. (2011). Rank - 1 / 2: A simple way to improve the ols estimation of tail exponents. *Journal of Business & Economic Statistics*, 29(1):24–39.

Garbinti, B., Goupille-Lebret, J., and Piketty, T. (2018). Income inequality in france, 1900–2014: Evidence from distributional national accounts (DINA). *Journal of Public Economics*, 162:63–77.

Garbinti, B., Goupille-Lebret, J., and Piketty, T. (2020). Accounting for Wealth-Inequality Dynamics: Methods, Estimates, and Simulations for France. *Journal of the European Economic Association*.

Guiso, L., Paiella, M., and Visco, I. (2005). Do capital gains affect consumption? estimates of wealth effects from italian households' behavior. *Long-run Growth and Short-run Stabilization: Essays in Memory of Albert Ando*.

Haziza, D., Beaumont, J.-F., et al. (2017). Construction of weights in surveys: A review. *Statistical Science*, 32(2):206–226.

Jenkins, S. P. (2016). Pareto models, top incomes and recent trends in UK income inequality. *Economica*, 84(334):261–289.

Kennickell, A. (2008). The role of over-sampling of the wealthy in the survey of consumer finances. *Irving Fisher Committee Bulletin*, 28.

Kennickell, A. (2019). The tail that wags: differences in effective right tail coverage and estimates of wealth inequality. *The Journal of Economic Inequality*.

Little, R. J. and Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology*.

Michelangeli, V. and Rampazzi, C. (2016). Indicators of financial vulnerability: a household level study. Questioni di Economia e Finanza (Occasional Papers) 369, Bank of Italy, Economic Research and International Relations Area.

Mumtaz, H. and Theophilopoulou, A. (2020). Monetary policy and wealth inequality over the great recession in the UK. an empirical analysis. *European Economic Review*, 130:103598.

Paiella, M. (2007). Does wealth affect consumption? Evidence for Italy. *Journal of Macroeconomics*, 29(1):189–205.

Ranalli, M. G. and Neri, A. (2011). To misreport or not to report?, The case of the Italian survey on household income and wealth. *Statistics in Transition new series*, 12(2):281–300.

Saez, E. and Zucman, G. (2016). Wealth inequality in the united states since 1913: Evidence from capitalized income tax data*. *The Quarterly Journal of Economics*, 131:qjw004.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, page 99.

Särndal, C.-E. and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. John Wiley & Sons, Ltd.

Schröder, C., Bartels, C., Grabka, M. M., König, J., Kroh, M., and Siegers, R. (2019). A novel sampling strategy for surveying high net-worth individuals—a pretest application using the socio-economic panel. *Review of Income and Wealth*.

Vermeulen, P. (2016). Estimating the top tail of the wealth distribution. *American Economic Review*, 106(5):646–50.

Vermeulen, P. (2018). How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, 64(2):357–387.

Waltl, S. R. (2021). Wealth inequality: A hybrid approach toward multidimensional distributional national accounts in europe. *Review of Income and Wealth*.

Yang, G. L. (1978). Estimation of a biometric function. *The Annals of Statistics*, 6(1):112–116.

8 Appendix: Tables and figures

TABLE 1
The missing gap: Pareto adjustments for gross wealth

	Coverage Ratios			Estimated tail households		
	Base	Adjusted	Implied	Total	95% C.I.	Missing
Country	(1)	(2)	(3)	(4)	(5)	(6)
IT	0.553	0.662	0.801	1,095,932	$\pm$ 4009.422	27,395.55
FR	0.673	0.772	0.806	1,145,012	$\pm$ 190.461	473.00
DE	0.827	0.870	0.997	1,419,375	$\pm$ 3016.263	7399.215
FI	0.917	1.035	1.061	87,290	$\pm$ 145.599	127.250
IT ($w_0 = 500k$)	0.553	0.685	0.787	2,972,082	$\pm$ 9865.057	23,799.14

Notes: Coverage ratios and estimated number of households in the tail. Re-weighting achieved with the Pareto-calibration method, using the calibration benchmarks from Equation 10.

TABLE 2
Simultaneous approach: final multivariate calibration adjustment factors

	Percentile									
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
Country	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IT	1.278	1.742	2.098	1.209	1.093	1.083	1.116	1.139	1.181	1.288
	(0.289)	(0.295)	(0.274)	(0.102)	(0.024)	(0.018)	(0.019)	(0.025)	(0.022)	(0.088)
FR	1.037	1.587	2.403	1.778	1.228	1.131	1.122	1.125	1.132	1.215
	(0.107)	(0.241)	(0.162)	(0.351)	(0.059)	(0.007)	(0.005)	(0.007)	(0.009)	(0.132)
DE	1.162	1.368	1.703	1.876	1.713	1.406	1.212	1.144	1.134	1.121
	(0.169)	(0.126)	(0.076)	(0.048)	(0.080)	(0.095)	(0.028)	(0.011)	(0.009)	(0.020)
FI	0.942	1.078	1.334	1.101	1.040	1.027	1.032	1.039	1.047	1.047
	(0.087)	(0.095)	(0.075)	(0.047)	(0.005)	(0.003)	(0.005)	(0.002)	(0.003)	(0.029)
IT ($500k$)	1.271	1.578	1.632	1.132	1.072	1.072	1.108	1.123	1.169	1.292
	(0.136)	(0.154)	(0.225)	(0.049)	(0.017)	(0.016)	(0.014)	(0.025)	(0.020)	(0.114)

Notes: Mean and coefficient of variation of overall adjustment factors a_i, Equations (11) and (12), from the multivariate calibration approach for imputation as a function of gross wealth percentiles.

TABLE 3
The missing gap: bootstrap results

Country	Top 1% (1)	Top 5% (2)	Top 10% (3)	Top 20% (4)	Bot 50% (5)	Gini (6)	Tail α (7)	S.r. (8)
	Wealth Shares							
	Base Survey							
IT	0.112	0.292	0.423	0.597	0.103	0.597	1.860	–
	(0.074)	(0.032)	(0.021)	(0.013)	(0.039)	(0.012)	(0.075)	
FR	0.168	0.352	0.485	0.655	0.073	0.655	1.768	–
	(0.089)	(0.036)	(0.022)	(0.012)	(0.040)	(0.011)	(0.052)	
DE	0.219	0.439	0.573	0.742	0.030	0.739	1.575	–
	(0.111)	(0.050)	(0.031)	(0.015)	(0.073)	(0.014)	(0.159)	
FI	0.120	0.285	0.416	0.595	0.101	0.596	2.162	–
	(0.050)	(0.019)	(0.012)	(0.007)	(0.025)	(0.007)	(0.069)	
	Survey & missing tail							
IT	0.284	0.431	0.537	0.677	0.083	0.676	1.370	1.000
	(0.020)	(0.017)	(0.015)	(0.012)	(0.004)	(0.011)	(0.016)	
FR	0.216	0.391	0.518	0.679	0.066	0.679	1.428	1.000
	(0.019)	(0.015)	(0.012)	(0.009)	(0.003)	(0.009)	(0.014)	
DE	0.326	0.517	0.632	0.778	0.026	0.775	1.348	0.981
	(0.019)	(0.014)	(0.012)	(0.008)	(0.002)	(0.007)	(0.019)	
FI	0.144	0.305	0.432	0.606	0.098	0.608	1.632	1.000
	(0.007)	(0.006)	(0.005)	(0.004)	(0.003)	(0.004)	(0.021)	
	Pareto-calibration & missing tail							
IT	0.323	0.482	0.584	0.712	0.075	0.707	1.431	0.996
	(0.016)	(0.016)	(0.016)	(0.012)	(0.004)	(0.011)	(0.016)	
FR	0.292	0.455	0.571	0.715	0.059	0.713	1.422	1.000
	(0.010)	(0.009)	(0.008)	(0.006)	(0.002)	(0.006)	(0.014)	
DE	0.345	0.537	0.648	0.787	0.025	0.784	1.358	0.952
	(0.019)	(0.015)	(0.013)	(0.009)	(0.002)	(0.008)	(0.019)	
FI	0.205	0.363	0.483	0.643	0.089	0.642	1.668	0.993
	(0.008)	(0.008)	(0.007)	(0.006)	(0.002)	(0.006)	(0.021)	
	Par-cal, Proportional allocation & missing tail							
IT	0.342	0.498	0.597	0.721	0.073	0.716	1.475	0.996
	(0.026)	(0.022)	(0.020)	(0.015)	(0.005)	(0.014)	(0.011)	
FR	0.295	0.458	0.572	0.716	0.058	0.715	1.448	1.000
	(0.011)	(0.009)	(0.008)	(0.006)	(0.002)	(0.006)	(0.019)	
DE	0.342	0.535	0.647	0.786	0.025	0.784	1.359	0.952
	(0.022)	(0.017)	(0.014)	(0.009)	(0.002)	(0.008)	(0.012)	
FI	0.209	0.367	0.486	0.645	0.089	0.644	1.649	0.993
	(0.009)	(0.009)	(0.008)	(0.006)	(0.003)	(0.006)	(0.019)	

TABLE 3
The missing gap: bootstrap results

	Wealth Shares							
	Top 1%	Top 5%	Top 10%	Top 20%	Bot 50%	Gini	Tail α	S.r.
Country	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Par-cal, Single-iteration approach & missing tail							
IT	0.400	0.556	0.649	0.762	0.064	0.712	1.370	0.996
	(0.123)	(0.096)	(0.077)	(0.052)	(0.014)	(0.052)	(0.016)	
FR	0.297	0.463	0.576	0.715	0.070	0.705	1.428	0.998
	(0.025)	(0.020)	(0.016)	(0.012)	(0.004)	(0.015)	(0.014)	
DE	0.348	0.524	0.634	0.772	0.035	0.759	1.348	0.941
	(0.059)	(0.043)	(0.034)	(0.021)	(0.004)	(0.028)	(0.019)	
FI	0.202	0.362	0.481	0.642	0.093	0.622	1.632	0.989
	(0.040)	(0.033)	(0.027)	(0.019)	(0.005)	(0.010)	(0.021)	
	Simultaneous approach							
IT	0.290	0.465	0.580	0.716	0.074	0.701	1.475	0.996
	(0.019)	(0.017)	(0.016)	(0.013)	(0.005)	(0.011)	(0.019)	
FR	0.295	0.461	0.575	0.714	0.071	0.705	1.432	1.000
	(0.016)	(0.014)	(0.013)	(0.010)	(0.003)	(0.006)	(0.017)	
DE	0.323	0.506	0.621	0.764	0.037	0.759	1.378	0.940
	(0.016)	(0.012)	(0.012)	(0.009)	(0.003)	(0.009)	(0.015)	
FI	0.174	0.344	0.476	0.647	0.087	0.640	1.767	0.992
	(0.008)	(0.008)	(0.008)	(0.007)	(0.003)	(0.006)	(0.029)	

Notes: Standard deviation reported in parenthesis. Wealth share by percentile, Gini inequality coefficients and Pareto tail parameters for Italy, France, Germany and Finland, estimated using different adjustments for the HFCS data, and accounting for the unobserved part of the Pareto tail. Survey weights are replaced by bootstrap weights. Threshold set at net wealth of EUR 1 million. Success rates (S.r.) report the observed probability of convergence for calibration.

TABLE 4
The missing gap: bootstrap results with 500k threshold (IT)

	Wealth Shares							
	Top 1%	Top 5%	Top 10%	Top 20%	Bot 50%	Gini	Tail α	S.r.
Country	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Base Survey				
IT	0.112	0.292	0.423	0.597	0.103	0.597	1.860	–
	(0.074)	(0.032)	(0.021)	(0.013)	(0.039)	(0.012)	(0.075)	
				Survey & missing tail				
IT	0.276	0.426	0.533	0.674	0.084	0.674	1.459	1.000
	(0.016)	(0.014)	(0.011)	(0.009)	(0.003)	(0.008)	(0.008)	
				Pareto-calibration & missing tail				
IT	0.308	0.473	0.577	0.708	0.078	0.701	1.429	1.000
	(0.014)	(0.008)	(0.006)	(0.005)	(0.003)	(0.005)	(0.008)	
				Par-cal, Proportional allocation & missing tail				
IT	0.316	0.479	0.582	0.711	0.077	0.705	1.487	1.000
	(0.018)	(0.011)	(0.008)	(0.006)	(0.003)	(0.006)	(0.007)	
				Par-cal, Single-iteration approach & missing tail				
IT	0.362	0.528	0.627	0.747	0.069	0.701	1.429	0.989
	(0.054)	(0.044)	(0.035)	(0.024)	(0.007)	(0.031)	(0.008)	
				Simultaneous approach				
IT	0.292	0.471	0.581	0.716	0.077	0.700	1.489	0.989
	(0.021)	(0.015)	(0.013)	(0.011)	(0.005)	(0.011)	(0.009)	

Notes: Standard deviation reported in parenthesis. Wealth share by percentile, Gini inequality coefficients and Pareto tail parameters for Italy estimated using different adjustments for the HFCS data, and accounting for the unobserved part of the Pareto tail. Survey weights are replaced by bootstrap weights. Threshold set at net wealth of EUR 500,000. Success rates (S.r.) report the observed probability of convergence for calibration.

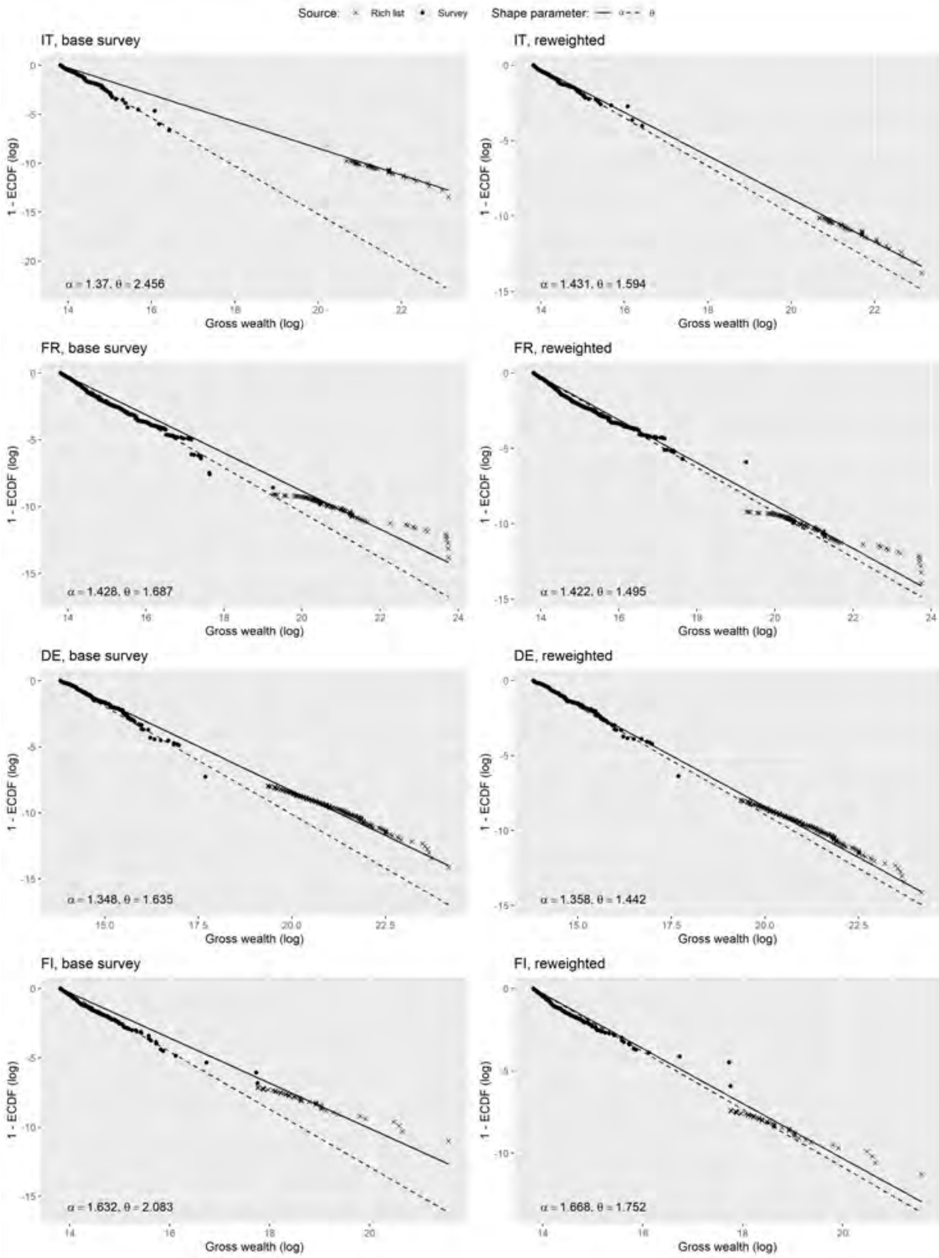
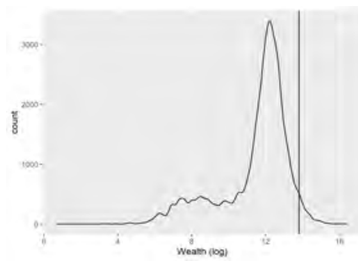
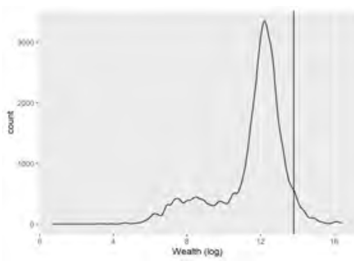


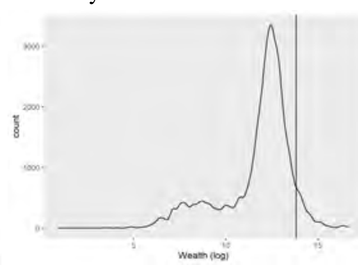
FIGURE 1: Pareto Tail Re-weighting. Empirical cumulative distribution functions (log scale) for survey wealth distributions in the Pareto Tail. Re-weighting achieved by using the Pareto-calibration method, using the calibration benchmarks from Equation (10). θ parameters estimated using survey data only, α estimated using Vermeulen's 2018 regression method with imputed rich list.



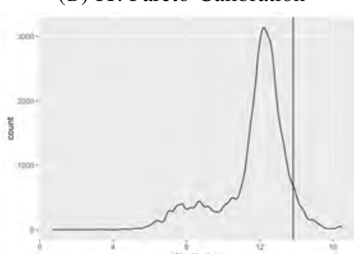
(A) IT: Base Survey



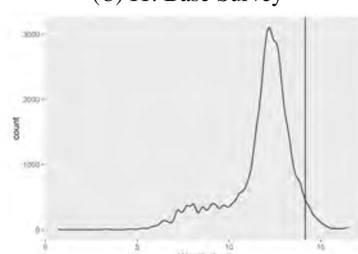
(B) IT: Pareto-Calibration



(C) IT: Base Survey

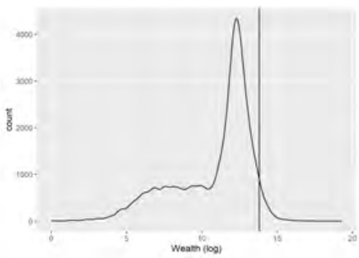


(D) IT: Pareto-Calibration

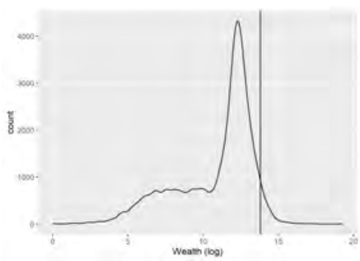


(E) IT: Pareto-Calibration

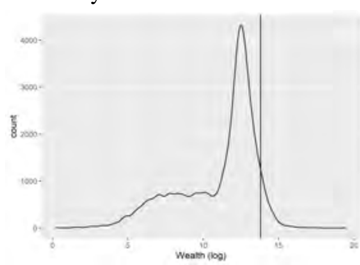
FIGURE 2: Italy: Empirical probability densities before (a) and after (b) Pareto-calibration, (c) Pareto-calibration & proportional allocation, (d) Single-iteration approach, and (e) Simultaneous approach.



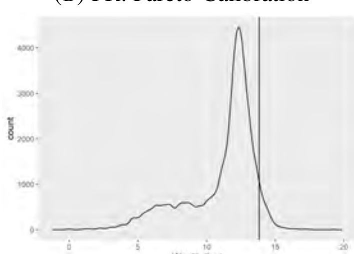
(A) FR: Base Survey



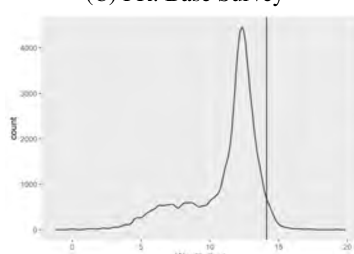
(B) FR: Pareto-Calibration



(C) FR: Base Survey

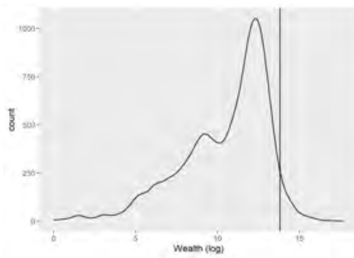


(D) FR: Pareto-Calibration

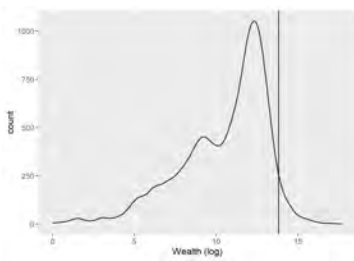


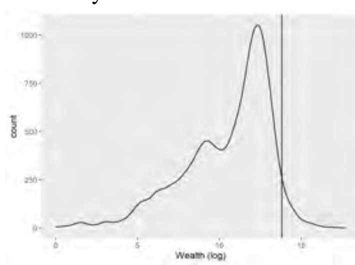
(E) FR: Pareto-Calibration

FIGURE 3: France: Empirical probability densities before (a) and after (b) Pareto-calibration, (c) Pareto-calibration & proportional allocation, (d) Single-iteration approach, and (e) Simultaneous approach.



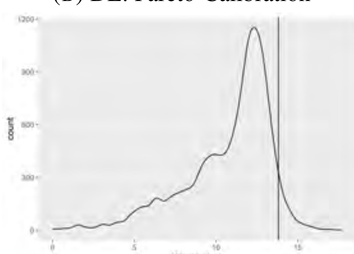
(A) DE: Base Survey

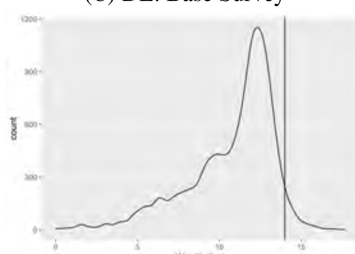




(B) DE: Pareto-Calibration

(C) DE: Base Survey

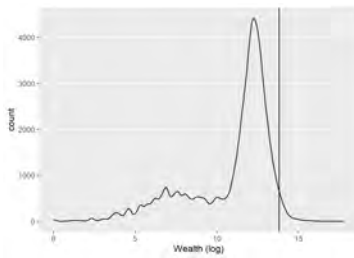




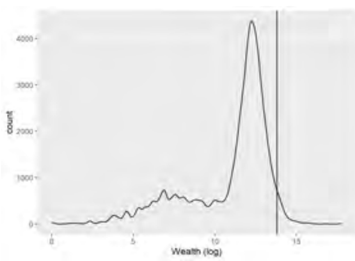
(D) DE: Pareto-Calibration

(E) DE: Pareto-Calibration

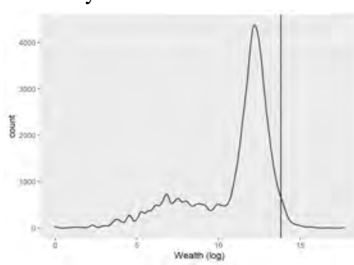
FIGURE 4: Germany: Empirical probability densities before (a) and after (b) Pareto-calibration, (c) Pareto-calibration & proportional allocation, (d) Single-iteration approach, and (e) Simultaneous approach.



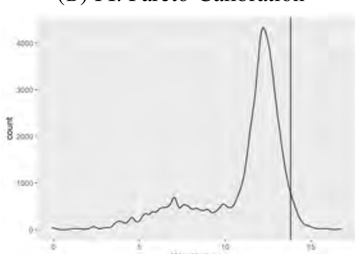
(A) FI: Base Survey



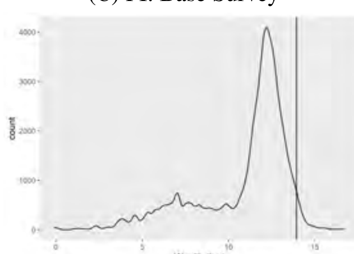
(B) FI: Pareto-Calibration



(C) FI: Base Survey

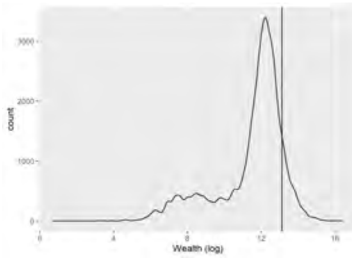


(D) FI: Pareto-Calibration

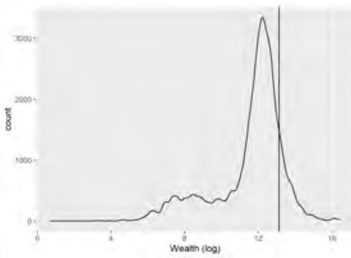


(E) FI: Pareto-Calibration

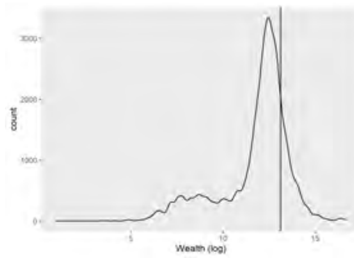
FIGURE 5: Finland: Empirical probability densities before (a) and after (b) Pareto-calibration, (c) Pareto-calibration & proportional allocation, (d) Single-iteration approach, and (e) Simultaneous approach.



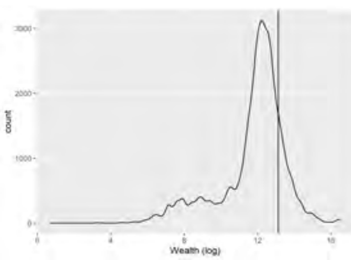
(A) IT: Base Survey



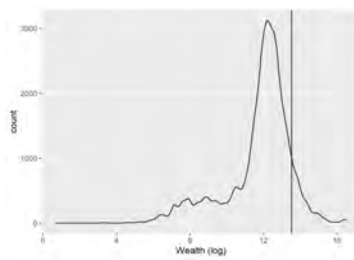
(B) IT: Pareto-Calibration



(C) IT: Base Survey



(D) IT: Pareto-Calibration



(E) IT: Pareto-Calibration

FIGURE 6: Italy: Empirical probability densities before (a) and after (b) Pareto-calibration, (c) Pareto-calibration & proportional allocation, (d) Single-iteration approach, and (e) Simultaneous approach. Threshold set at EUR 500,000.

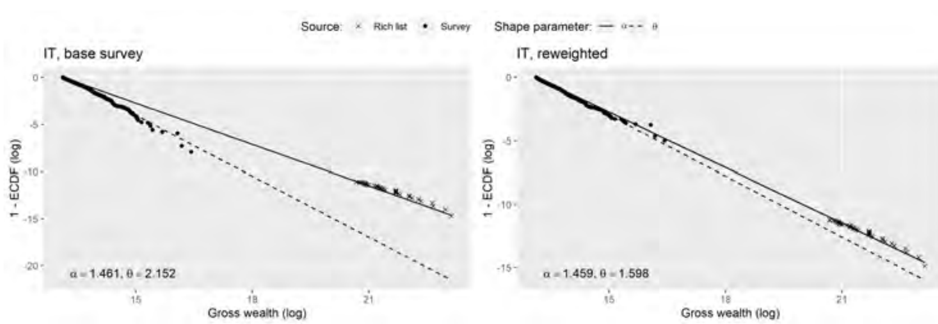


FIGURE 7: Pareto Tail Re-weighting. Empirical cumulative distribution functions (log scale) for survey wealth distributions in the Pareto Tail for Italy. Re-weighting achieved by using the Pareto-calibration method, using the calibration benchmarks from Equation (10). Threshold set at net wealth of EUR 500,000. θ parameters estimated using survey data only, α estimated using Vermeulen's 2018 regression method with imputed rich list.