BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Bridging techniques in the redesign of the Italian Survey on Household Income and Wealth

by Romina Gambacorta and Eleonora Porreca

Number 719

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Bridging techniques in the redesign of the Italian Survey on Household Income and Wealth

by Romina Gambacorta and Eleonora Porreca

Number 719 – October 2022

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

BRIDGING TECHNIQUES IN THE REDESIGN OF THE ITALIAN SURVEY ON HOUSEHOLD INCOME AND WEALTH

by Romina Gambacorta* and Eleonora Porreca*

Abstract

The design of the Bank of Italy's Survey on Household Income and Wealth was revised in 2020 to reduce non-sampling error in households' income and wealth and improve data quality. The new sample allocation resulted in greater participation in the upper parts of the income distribution, determining a reduction in standard errors and in the bias of income and wealth estimators. However, the revision of the sample makes it difficult to compare the results with those obtained in previous survey waves. This paper discusses different weighting systems for taking these differences into account, obtained following three main methodological approaches: cell weighting, raking and inverse probability weighting. Comparing results across different dimensions, the method that produces the most reliable results is based on the use of the raking technique and, therefore, it is the one recommended for time series analysis.

JEL Classification: C83, D31.
Keywords: survey sampling design, weighting, calibration, income and wealth distribution.
DOI: 10.32057/0.QEF.2022.0719

Contents

* Bank of Italy, Directorate General for Economics, Statistics and Research.

1. Introduction[1]

One of the main problems in household surveys is measurement bias in income and wealth due to the non-participation of well-endowed households (Sànchez Muñoz, 2011, Vermeulen, 2018). This problem also affected the Survey on Households Income and Wealth (SHIW) conducted by the Bank of Italy since the '60s (D'Alessio and Faiella, 2002), leading to the underestimation of households' financial assets (D'Aurizio et al. 2006, Neri and Ranalli, 2012), income sources (Neri and Zizza, 2010) and generally producing estimates inconsistent with macro aggregates (D'Alessio and Neri, 2015). Finally, as this bias could be attributable – at least in part – to the lack of richer households in the sample, this also conveys effects in the distributional analysis and possibly inaccurate results of poverty or inequality indexes.

To overcome this problem, many household surveys have adopted different sampling strategies for covering rich households. Within the Household Finance and Consumption Survey (HFCS), the Euro area harmonised national central banks' households' survey, in 2017 over three-quarters of the countries used a strategy to sample affluent households and, although with different methods, they improved efficacy (HFCN, 2020). Due to the lack of reliable external sources, Italy was one of the few countries not adopting any sampling strategy to contrast selective non-response.

Thanks to a collaboration between the Bank of Italy and the Italian National Statistical Institute, it was possible to redesign the 2020 wave of the SHIW. Specifically, using administrative register data on the income it was feasible to add an optimal stratification of households within the sampled municipalities to the second stage of the design. Considering the increasing use of the SHIW also for the analysis of households' financial vulnerability, the 2020 sample was further complemented with an additional sample of indebted households (2.000 households), stratified with respect to their level of outstanding debt recorded in the administrative central credit register, to improve estimates on households' indebtedness. If on one hand this new design allows to better represent the distribution of income, wealth, and household indebtedness, thus reducing standard errors and the measurement bias due to selective non-participation, on the other hand it introduces a structural break in the data collected. The new sample allocation resulted in greater participation of the households in the upper part of the income and debt distribution. Therefore, the estimates for these variables were revised upwards,

[1] The authors wish to thank Giovanni D'Alessio, Silvia Fabiani and Andrea Neri for their suggestions and advice. The views and opinions expressed are those of the authors and do not necessarily reflect those of the Bank of Italy.

as have their variance and inequality. This discrepancy created difficulties in comparing the new data with those collected in the previous waves.

Two main approaches are possible to account for the revision of the design and allow data comparability over time. One is the correction of past data to account for the bias due to the lower presence of observations in the top tail of the distribution, and the other consists in adjusting newly collected data leading them back to what one would have obtained by adopting the previous sample design. The first method is not applicable in the case of the SHIW, as it would need administrative data and a certain number of observations in the top tail of the distribution of income and debt, which are both missing in past editions of the survey (going back to 1989, the starting date of the historical series). In the absence of such data, this approach would rely only on heavy assumptions with respect to the top tail of the distribution of households' income and debt, which would be difficult to verify. For these reasons, in this paper, we focus on the second approach.

The literature provides several techniques to estimate the effects of survey changes (see van den Brakel et al. 2020 for a review). In general, when the data collection phase is modified (questionnaire design, mode of interview, field strategy) leading to inconsistent information between the old and the new scheme, parallel data collections should be implemented. This, under reasonable conditions such as sufficient sample size in the treatment groups, would allow estimating the impact of survey discontinuities by comparing the treatment groups, but it is extremely costly and cannot always be implemented. In the case of the SHIW, the share of the sample that remained unaffected by the redesign was represented by the panel component. Unfortunately, due to the Covid-19 pandemic that dramatically affected households' participation, the final sample size of the survey was about half of the planned. The reduced size of the panel component weakened the use of this component for redesign effect estimates.

When parallel data collection is not feasible, other strategies can be applied. For example, when the questionnaire or data classification is modified, a multiple imputation approach can be adopted (e.g. Rothbaum 2017), which imputes information gathered under the new scheme to the traditional sample treating the problem as missing values. Also, time series models can be adopted that decompose the observed series into several components, such as trend, seasonal and survey discontinuity components. However, the estimation of the impact of the changes is impossible when the new survey edition has one observation only. Moreover, as new observations are added to the series, the effect of discontinuities might change and estimates revisions should be implemented. Time series models are not suitable also in cases when survey redesigns coincide with remarkable real changes in the observed phenomenon, as in the case of the SHIW redesign during the Covid-19 pandemic.

The choice of methods for estimating survey discontinuities strictly depends on the type and the time of the redesign, the budget and the timeliness of publication requirements. This paper complements this literature by proposing weighting strategies that can be used to compare estimates in different domains between the old and the new survey scheme in cases where the existing approaches are not suitable. The main contribution of this paper is to achieve this result by adopting reweighting strategies, usually used in the literature in the framework of adjustments for non-response or to study the causal treatment effect. Particularly, we compare three different strategies in reweighting collected data, namely cell weighting, raking, and inverse probability weighting, to assign to households in the new sample the same probability of selection they would have had in the old sampling strategy.

The paper is organized as follows. In section 2 the SHIW new survey design is described. Section 3 illustrates the three reweighting methodologies adopted. Section 4 discusses the results also with respect to the distribution of income and wealth using different reweighting approaches and proposes the strategy to be used in official statistics to compare the last two waves. In the last section, the main conclusions are drawn.

2. The redesign of the Italian Survey on Household Income and Wealth

The Bank of Italy SHIW collects information on the households' socio-demographic and economic conditions since the '60s adopting a two-stage stratified sample design[2]. In the first stage, a stratified sample of about 400 municipalities (Primary Sampling Units, PSUs) is selected. Municipalities are stratified by region and demographic size; the largest municipalities (with more than 40,000 inhabitants) are always selected (self-representing units SRUs), while the others are selected with a probability that grows proportionally with their size. In the second stage, within the selected municipalities, households (Secondary Sampling Units SSUs) are selected randomly from the municipality registers.

Up to 2016, due to the lack of information about the economic conditions of selected units at the design stage, non-respondent households were replaced by other units randomly selected among those living within the same municipality. This approach did not take into account that the

[2] The design was revised in 1989 to increase the sample size to 8,000 households and to include a panel component which represents about half of the sample. Panel households exit from the sample only in case of refusal or incurred ineligibility and, to account for panel attrition, in every edition of the survey a refresh component is drawn from households who have participated for the first time in the previous survey waves.

participation rate tends to decrease with income and wealth (D'Alessio and Faiella, 2002) and that the very affluent households are hard to reach and difficult to persuade to participate in the survey, being severely under-represented in the final sample. Those households, although small in number, detain a large share of both income and wealth. Their absence in the sample determines therefore a large underestimation of the main target variable of the survey such as income, wealth and their components (Neri and Ranalli, 2011; D'Alessio and Neri, 2015).

In 2020, thanks to the collaboration between the Bank of Italy and the Italian National Statistical Institute, it was possible to merge administrative data from the Italian Tax Register to the Italian Population Register used to draw the sample.

These lists, although very useful, present some drawbacks. First, there are differences in their definitions with respect to those adopted in the survey data (for example in the definition of household in the population register). In the Tax Register some units can be missing, as only individuals with income above a specific threshold are observed. Furthermore, register data are only available with a two-year time lag, so it is not possible to use updated information to draw the sample. Finally, quality issues may affect official records, mainly those due to tax evasion or tax elusion.

To overcome these problems while retaining the informative content of administrative data, Barcaroli et al. (2021) propose an optimal sampling strategy where variables from the tax records are considered as proxies of the households' economic situation. The optimal sample was selected using a genetic algorithm, which jointly selects the strata boundaries and the sample allocation for a multivariate population that minimizes the total cost of the sample given the precision constraints (Ballin and Barcaroli, 2013).

In practice, as register data are considered as proxies of the target variables (total income, dependent employment income, self-employment income, pension income, rents), their goodness of fit was used to calculate the inflation term of the population estimates of the variance in the strata. The precision constraint was set as the expected coefficient of variation of the target means estimates in the main Italian geographical areas (five groups of regions NUTS1) to be less or equal to 5 per cent. The methodology was used in the second stage of the 2020 SHIW design to select non-panel households, leaving the first stage and the panel units at the second stage (about 4,000 households) unchanged[3]. The algorithm selected an optimal sample with 10 strata within each geographical area and a total sample size of about 6.000 households. Simulations show that

[3] This choice, although sub-optimal, was taken to contain revision in the organisational and fieldwork procedures.

the new design was effective in reducing both the sampling variance and the estimation bias due to non-response.

Finally, considering the increasing use of the SHIW also for the analysis of households' financial vulnerability, the 2020 SHIW sample was further complemented with an additional sample of indebted households (2,000 households) stratified with respect to their level of outstanding debt (in five strata for bad debt and non-bad debt as resulting from the central credit register) to improve estimates on households' indebtedness.[4] The size of the total sample for the survey was therefore set to 12,000 households.

Unfortunately, the Covid-19 pandemic strongly affected households' participation in the survey. To cope with this situation, the fieldwork of the survey has been shifted forward by one year, and specific strategies have been adopted to reduce non-response due to the fear of contagion (such as allowing for a share of telephone interviews). Nevertheless, the 2020 final sample size was much smaller than predicted, with a total of 6,239 households (of which about 3,000 panel households). The new sample design made it possible to have a sample of all income and debt strata, capable of producing more efficient estimates than those that would have been obtained with the traditional design. Table 1 compares the households' sample distribution across income and debt strata obtained in 2016 and 2020. The main effect of the new design concerns the income distribution of the final sample. The share of households in the 2016 sample belonging to the highest income strata is close to zero while in the 2020 sample the higher strata are more equally distributed. Accordingly, in the 2020 sample the share of households belonging to the poorest income stratum is lower than in 2016. Furthermore, due to the additional sample of indebted households, the 2020 wave better captures the households' indebtedness: 84 per cent of households reported not having any debt in 2016, versus 62 per cent in 2020. The share of households in the highest debt stratum of both bad and non-bad debt is remarkably higher in 2020 than in 2016. It should be noted that also this further sample has the effect to complement the sample with wealthier families as those are the ones that possess the necessary guarantees to reach the highest levels of indebtedness.

[4] See di Salvatore et al. (2020) for a detailed description of the indebted households sample.

Table 1: households' distribution of the SHIW sample across income and debt strata

Income strata[1]	Frequency		Percent	
	2016	**2020**	**2016**	**2020**
1	6,235	3,882	84.03	62.22
2	796	850	10.73	13.62
3	287	360	3.87	5.77
4	36	156	0.49	2.50
5	27	149	0.36	2.39
6	17	146	0.23	2.34
7	6	225	0.08	3.61
8	13	210	0.18	3.37
9	2	130	0.03	2.08
10	1	131	0.01	2.10
Total	7,420	6,239	100.00	100.00
Debt strata	**Frequency**		**Percent**	
	2016	**2020**	**2016**	**2020**
Not indebted	6,256	4,081	84.31	65.41
Non-bad debt				
1	84	221	1.13	3.54
2	159	228	2.14	3.65
3	165	261	2.22	4.18
4	177	298	2.39	4.78
5	286	843	3.85	13.51
Bad debt				
1	140	101	1.89	1.62
2	33	19	0.44	0.30
3	37	26	0.50	0.42
4	28	34	0.38	0.54
5	55	127	0.74	2.04
Total	7,420	6,239	100.00	100.00

Source: Authors' elaborations on SHIW 2016 and 2020 data. 1: income strata aggregated by geographic area. Frequencies and percentages of households are not weighted.

The major problem with the new sampling design resides in the difficulty of comparing the estimates obtained according to the new design with those from the previous editions of the survey. This aspect is particularly important for a survey with long historical data series such as the SHIW. At the design stage, it was thought to overcome this difficulty by resorting to the panel component, which had to be composed of 4,000 units. However, due to the pandemic, the number of panel families has been reduced to less than 3,000 households, which narrows the analysis to a numerically restricted subsample making comparisons less reliable. Moreover, the pandemic may have had different effects in terms of non-response among the different population groups, potentially weakening the effectiveness of these comparisons. For these reasons, in this paper, we show some alternative techniques to compare the results of the 2020 survey with those of the previous edition conducted in 2016, which can be used as an alternative to the comparison to the panel component. The techniques are based on the use of different data re-weighting techniques that aim to ensure that the 2020 sample is as similar as possible to what it would have been in the absence of a revision of the design.

3. Methods

To produce estimates from SHIW 2020 comparable with the previous waves, i.e. net of the sampling redesign, we modify the weighting process used for the 2020 wave.

Considering the sampling design of the SHIW and a variable of interest y, an unbiased estimator of the population mean is given by the Horwitz-Thompson estimator (Kish, 1965):

$$\bar{y} = \frac{\sum_h \sum_i \sum_j y_{hikj} w_{hikj}}{\sum_h \sum_i \sum_j w_{hikj}} \quad j = 1, \dots, n_{hik} \quad k = 1, \dots, K \; i = 1, \dots, a_h \quad h = 1, \dots, H$$

where y_{hikj} is the value of y observed for household j belonging to second-stage stratum *k*, residing in municipality i belonging to first-stage stratum h, while w_{hikj} is its sampling weight.

The weighting process for the 2020 wave consists of the following major steps (in what follows, we will refer to weights obtained with this procedure as the standard 2020 weights):

i. The initial weight $w_{hik}^{(0)}$, defined as the design weight given to each household, is computed as the product of the inverse of the probability of selecting municipality *i* in the stratum h (first stage design weights) and the inverse of inclusion probability of a household belonging to stratum k (second stage design weights)[5] and residing in municipality i in stratum h:

$$w_{hik}^{(0)} = \left(\frac{1}{m_h} \frac{P_h}{P_{hi}}\right) \frac{N_{hik}}{n'_{hik}} \qquad h = 1, \dots, H \; i = 1, \dots, a_h \; k = 1, \dots, K$$

where P_h e m_h are, respectively, the resident population and the number of sample municipalities in the h^{th} first stage stratum, P_{hi}, is the resident population in the i^{th} municipality of stratum h, while N_{hik} and n'_{hik} are, respectively, the number of households residing and selected (theoretical sample) in the i^{th} municipality on the h^{th} first stage stratum and belonging to the k^{th} second stage stratum;

ii. $w_{hik}^{(0)}$ is adjusted for the selection due to the unit non-response by multiplying it by the inverse of response rate in stratum *k* to which each household belongs, obtaining $w_k^{(1)}$:

[5] Strata used for the construction of second stage design weights are collapsed compared to those used for the selection of the sample, to reduce the variability of the final weights. In particular, households are grouped into 6 strata constructed on family income and 3 strata defined by their level of indebtedness.

$$w_k^{(1)} = w_{hik}^{(0)} \frac{n'_k}{n_k}$$

where n'_k e n_k are, respectively, the total number of selected households (theoretical sample) and the total number of respondents (final sample) in the k^{th} second stage stratum;

iii. $w_k^{(1)}$ is modified to account for the attrition in the panel component, obtaining $w^{(2)}$, by post-stratifying panel households to the distributions of some characteristics measured in the whole sample in the previous wave and by adjusting to reproduce the optimal share of the panel component (estimated at about 50 per cent of the sample)[6];

iv. finally, the weights are post-stratified using external information correlated with the core economic variables to improve the accuracy of estimators. In particular, weights are modified to reproduce the same characteristics as the population in terms of sex, age (7 classes), geographic area (3 classes), size of the municipality (4 classes), education (2 classes) and household size composition (5 classes), to obtain final weights $w_j^{(3)}$:

$$w_j^{(3)} = w_j^{(2)} \gamma_j$$

where γ_j is the post-stratification factor for the $j-th$ post-stratification cell for variable γ.[7]

To construct weights for the SHIW 2020 capable of filtering estimates from the effect of sampling redesign, a further step is added before the post-stratification (step iv). At this new step, the weights $w^{(2)}$ are adjusted such that the weighted distributions of households by their outstanding debt (from credit register) and their income (from Tax records) in 2020 are close to the one observed on the sample in 2016. However, to take into account the changes that occurred in the population between 2016 and 2020, we define the allocation of the 2016 sample across the strata using the same administrative sources and methodology adopted to construct strata in 2020.

After this process, the standard post-stratification step is implemented to make the demographic characteristics of the SHIW 2020 compliant with the population characteristics in the same reference year.

To conduct the additional weighting step, three different weighting approaches are considered: 1) cell weighting, 2) raking and 3) inverse probability weighting (IPW).

[6] For more details on the weighting process of the panel component in the SHIW see Faiella and Gambacorta (2007).

[7] At the end of the weighting process, after the post-stratification step, weights are trimmed at the 99-th percentile.

Cell weighting and raking are standard methods used in the sampling literature on calibration estimators to address post-stratification (Deville e Särndal, 1992) and non-response (Kalton 2003 Kalton and Kasprzyk 1986).

According to the cell weighting approach, the totals of the joint distribution of the new design variables in 2020 are adjusted to match those observed in 2016 on a cell-by-cell basis. In particular, the weights in each cell of the joint distributions are obtained as the ratio of the frequency observed in 2016 to the frequency observed in 2020 (weighted by $w^{(2)}$). Cell weighting assumes that households have the same probability of being interviewed in 2016 within each cell. The adjustment within cells might lead to increased weight variability, especially when the sample size in each cell is small and when it greatly differs with respect to the target distribution.

Raking is an iterative procedure that conforms the joint marginal distributions of debt and income strata in 2020 to those observed in 2016. The initial weights, $w^{(2)}$, within each cell are adjusted such that first the row totals in 2020 match the row totals in 2016, and then the same exercise is repeated for column totals until convergence is reached. Raking assumes that the probability of being interviewed in 2016 is equal for each household in a cell and that this probability only depends on the marginal distributions related to that cell. This additional assumption leads to less variability than cell weighting, especially when the number of cells is large and the sample size within cells is small. However, when this assumption is not met, the reduction bias of the estimators might be limited.

Inverse probability weighting (IPW) originates from the sampling literature, where it is widely used for handling unit non-response by adjusting the weight by the inverse of the response probability (Iannacchione et al. 1991, Kim and Kim 2007, Kim and Riddles 2012). This approach has also been adopted and extended in the literature on the causal treatment effect estimation and in particular in the context of the propensity score method. In this literature, the IPW is used to estimate, on top of the average treatment effect (ATE), the average treatment effect on the treated (ATT), where the average outcome of the control group is weighted to obtain the average outcome as if this group had received the treatment, i.e. the counterfactual average outcome. This method is also used in the literature simulating the wage distribution when some characteristics (such as age or education) were changed (e.g. DiNardo et al. 1996, Frolich 2007). These studies show that it is possible to derive, not only the average counterfactual outcome but also the counterfactual distribution of the outcome of interest. For example, DiNardo et al. (1996) estimate the counterfactual distribution of wages in 1988 if workers' union status had remained at its 1979 level; Frolich (2007) generalises the Blinder-Oxaca wage gap by estimating the counterfactual wage distribution of men if they had the human capital characteristics of women.

In a similar vein, we aim at deriving the counterfactual distribution of core variables in 2020 if the sampling design had not changed since 2016, i.e. if the allocation of the sample into the income and debt strata used for the new sampling design had remained at their 2016 levels, once the changes in the distribution of the population between the two waves have been taken into account. To this aim, strata in 2016 have been defined using the same methodology used in 2020 but referring to the 2016 population. Therefore, the initial weights, $w^{(2)}$ are multiplied by the IPW weight $\frac{e(X_j)}{1-e(X_j)}$, where $e(X_j)$ is the probability of a household j of being interviewed in 2016, i.e. the propensity score. This probability is estimated using a logit model and considering income and debt strata among the covariates X_j. Appendix A.1 shows how the IPW weight is derived in the context of the causal treatment effect.

The main advantage of this method is that it is more flexible with respect to the previous two, especially in the case of high dimensional X_j or when it includes continuous or interaction variables. However, it relies on unconfoundedness and overlapping assumptions.[8] The first one states that all the variables influencing the probability of being interviewed in 2016 should be observed, whereas the second ensures that households with the same value of covariates have a positive probability of being interviewed in 2016 and in 2020, i.e. it ensures comparability in terms of covariates between households interviewed in the two waves. Unconfoundedness is not testable and should be justified by the data quality. The overlapping assumption is testable by analysing the support of $e(X_j)$, which is very sensitive to the choice of covariates and misspecification. The accuracy of the propensity score can be evaluated by analysing the balance of X_j between households interviewed in 2016 and those interviewed in 2020.

To evaluate these strategies, we gather from the literature on the bias-variance trade-off adopted to evaluate non-response weight adjustments (Kish, 1992, Little, 1986). It should be noted that, in this context, our goal is not to assess the efficiency of the new sampling design, but to evaluate the redesign-adjusted weights with respect to the standard 2020 ones in terms of reproducing a sample as similar as possible to that we would have obtained in the case the design had not changed. Therefore, our measure of "bias" is not the difference between an estimate and its "true value", which we assume to be reduced by the new design, but is defined as a distance

[8] See Appendix A.1 for the formalization of the assumptions.

of the households' distributions across redesign strata between 2016 and 2020. Furthermore, we won't look at the changes in the variance of relevant estimates using a different set of weights, as we know that using weighing adjustments associated with the survey outcome can increase efficiency (Little and Vartivarian, 2005), so in principle, standard 2020 weights are more appropriate. In our case, we focus only on the increase in the weights' variability due to the further adjustment step to the weighting process, as we want to quantify which among the proposed methods minimises this potential loss of efficiency.

Following this approach, the different strategies proposed to minimise the redesign effect are compared in terms of underlying assumptions, reduction bias of redesign strata distributions with respect to 2016 and variance inflation. The latter is analysed using the efficiency loss function L_w (Kish 1965, 1992), which measures the percentage increase in the variance of an estimator due to the use of weights:[9]

$$L_w = \frac{V_w - V_{uw}}{V_{uw}} = cv_w^2$$

where V_{uw} and V_w are respectively the variances of the unweighted and weighted estimator, and cv_w is the coefficient of variation of weights.

To measure the impact on estimators' variability due to the additional adjustment step we calculate the percentage increase in the efficiency loss function of redesign adjusted weights with respect to the standard weighting process of the SHIW 2020:

$$\Delta L_{w^*} = \frac{L_{w^*} - L_w}{L_w} = \frac{cv_{w^*}^2 - cv_w^2}{cv_w^2}$$

where L_{w^*} and L_w are, respectively, the loss function of the redesign adjusted and standard weighted estimators, and cv_{w^*} and cv_w are the coefficients of variation of the redesign adjusted and standard weights.

To take into account the contemporaneous effect of both the bias reduction and efficiency loss, we also compare the distribution of redesign strata, assuming that the estimated distributions in 2016 are correct. In particular, the reduction in bias for each redesign variable is defined as:

$$bias = \sum_c \left| \hat{\theta}_{c,2020,w^*} - \hat{\theta}_{c,2016} \right|$$

[9] Although this formula is not exact in the case of stratified multistage surveys, it represents a good approximation of the true value in many cases (see Little et al., 1997).

where c indicates the $c-th$ category of each redesign variable, $\hat{\theta}_{2016}$ and $\hat{\theta}_{2020,w^*}$ are respectively the estimated relative frequencies in the $c-th$ category in 2016 (weighted by the standard SHIW weight) and in 2020 (weighted by the redesign adjusted weight).

We adopt the three weighting strategies using first a set of covariates, X_j, that includes only the new design variables (income and debt strata).[10] We call these weights baseline adjusted weights (B1, B2 and B3 referring to baseline adjusted weights using cell weighting, raking and IPW respectively). Then, we define compound adjusted weights obtained including further additional variables, correlated to the redesign variables (C1, C2 and C3 for compound adjusted weights using cell weighting, raking and IPW respectively). To identify these variables we have inspected the sample composition between the two waves along different dimensions related to income and wealth households' endowments. We find that, even after the adjustment of the baseline redesign weights, the distribution of the 2020 sample was still far from 2016 in terms of the share of income recipients within the households,[11] the share of employed workers in the sample, households' distribution across quartiles of property income (observed in 2016) and financial literacy.[12] According to external sources, these variables only have slightly changed during the observed period.[13] As a misalignment of the sample along these dimensions may end in a bias of estimates due to their strong correlation with target variables, we have added these four variables to post-stratification in the compound weights. In the following results, we present the compound version including the four mentioned additional variables. In Appendix A.2 we compare three compound versions, for each of the weighting strategies, starting from the most parsimonious specification and ending with the full (preferred) specification. These results show

[10] Income and debt strata are aggregated because of the very limited or null number of households in some categories observed in 2016. In particular, income strata are modified from ten income strata within each geographic region, to five income strata across all geographic regions, collapsing income strata from stratum five to ten into the fifth category. Debt strata are modified into six categories, where the first five are based on the outstanding debt level and the last one is obtained by aggregating the five bad debt strata into a unique bad debt category.

[11] This variable has been transformed into a dummy variable equal to one if the share of income earners within the household is more than half and equal to zero otherwise.

[12] Financial literacy is equal to zero when the household did not respond correctly to at least one of the three financial literacy questions (QTASSO, QINT, QRISK1), equal to one if the household provided one or two correct answers and equal to two if the household correctly answered to all the three questions.

[13] According to the National Institute for Statistics (ISTAT), the share of households with at least two employed persons remains quite stable between 2016 and 2020, i.e. it slightly decreases from 24.66 per cent to 24.38 per cent. Also according to the Italian Revenue Agency, the average number of income earners per family remains stable at 1.6 both in 2016 and in 2020. Referring to the share of the employee in the labour force, ISTAT shows that it increases by 0.89 percentage points between 2016 and 2020. Although very limited, we apply this variation to the 2016 distribution. Focusing on the quartiles of property income, which is a proxy for ownership, the EU-SILC survey shows that the share of households owning their main residence slightly decreases from 80.3 per cent in 2016 to 78.8 per cent in 2019 (the latest available wave). Finally, according to the Survey on the Financial Literacy of Italian Adults, conducted by the Bank of Italy, the overall financial literacy indicator did not change in 2020 with respect to the level estimated in 2017.

that increasing the number of additional variables does not increase weight variability, except for the cell weighting, and remarkably reduces the bias in most of the analysed dimensions.

4. Results

Table 2 shows that, as expected, the redesign adjustment increases the weight variability, independently of the weighting strategies adopted. Also adding further constraints to the baseline weights increases the efficiency loss by 23 per cent with cell weighting and by 11 per cent with raking and by 3 per cent with IPW with respect to the baseline weights.

Both among the baseline and the compound redesign adjusted weights, the strategy that guarantees the minimum efficiency loss with respect to the standard SHIW weights is raking. The number of observations with the IPW is lower than the total number of interviewed households because those excluded do not satisfy the common support assumption, i.e. those with a propensity score outside the common propensity score range formed by households interviewed in 2016 and 2020. Therefore, it is not possible to attribute a new weight to these households, representing a remarkable drawback of this approach. Raking leads to lower variability than cell weighting because in this context the number of cells is large (35 in the baseline specification and 1,680 in the compound specification) and the cell size is small. Therefore, in terms of underlying assumptions and efficiency loss, the most suitable approach is raking.

Table 2: weight distributions and efficiency loss

Year	Weight	Obs	Mean	SD	Min	Max	L_{w^*}	ΔL_{w^*}
2020	Standard SHIW weight	6,239	1	1.23	0.00	6.55	1.51	
Baseline redesign adjusted weights								
2020	B1 Cell weight	6,239	1	1.32	0.00	7.00	1.74	0.16
2020	B2 Raking	6,239	1	1.31	0.00	7.00	1.71	0.13
2020	B3 IPW	6,173	1	1.36	0.00	7.07	1.86	0.23
Compound redesign adjusted weights								
2020	C1 Cell weight	6,239	1	1.47	0.00	8.07	2.15	0.42
2020	C2 Raking	6,239	1	1.38	0.00	7.68	1.90	0.26
2020	C3 IPW	6,193	1	1.38	0.00	7.96	1.91	0.27

Source: Authors' elaborations on SHIW 2016 and 2020 data.

Table 3 shows the weighted distributions of redesign strata and additional variables used to obtain the compound weights. As expected, the standard SHIW weights lead to a dramatic difference in the distributions with respect to the 2016 wave, showing a sample composed of higher-income households, higher indebtedness, a lower share of employees vs. especially a higher share of self-employed, and a higher share of income earners within the household. Redesign adjusted weights remarkably decrease the bias, i.e. the differences between the two distributions. In particular, the compound version of the three strategies conforms the distributions of both redesign strata and additional variables to the SHIW 2016 much better than the baseline framework. Focusing on the compound version and on income strata, which is the most important redesign variable, raking minimises the bias much more than cell weighting and IPW. Turning to the additional variables, the highest reduction bias is not clear-cut across the three strategies: cell weighting and IPW seem to minimise the bias depending on the distribution considered, whereas raking reduction bias shows values very close to the minimum between the other two approaches.

Table 3: weighted distributions and reduction bias

Redesign variables								
		Income strata						
Year	Weight	1	2	3	4	5		Bias
2016	Standard SHIW weight	85.21	9.63	3.61	0.39	1.16		
2020	Standard SHIW weight	77.09	9.24	4.79	1.20	7.68		17.01
Baseline redesign adjusted weights								
2020	B1 Cell weight	88.86	6.93	3.14	0.26	0.80		7.30
2020	B2 Raking	87.45	7.87	3.42	0.35	0.92		4.47
2020	B3 IPW	89.30	7.06	2.77	0.23	0.64		8.18
Compound redesign adjusted weights								
2020	C1 Cell weight	87.39	8.48	2.91	0.30	0.92		4.37
2020	C2 Raking	86.14	9.21	3.36	0.31	0.99		1.87
2020	C3 IPW	87.62	8.32	2.91	0.31	0.84		4.82
		Debt strata						
Year	Weight	0	1	2	3	4	5	Bias
2016	Standard SHIW weight	81.32	3.00	3.41	3.06	3.32	5.88	
2020	Standard SHIW weight	69.53	4.49	3.75	4.52	5.42	12.29	23.59
Baseline redesign adjusted weights								
2020	B1 Cell weight	79.18	2.49	3.14	3.27	4.24	7.67	5.85
2020	B2 Raking	79.60	2.38	2.84	3.16	4.06	7.96	5.83
2020	B3 IPW	81.59	2.34	2.96	2.90	3.76	6.46	2.56
Compound redesign adjusted weights								
2020	C1 Cell weight	81.70	2.24	2.60	2.76	3.78	6.92	3.76
2020	C2 Raking	80.53	2.39	2.83	3.06	3.69	7.51	4.00
2020	C3 IPW	82.16	2.39	2.84	2.82	3.53	6.26	2.85
Additional variables								
		Property income class						
Year	Weight	1	2	3	4			Bias
2016	Standard SHIW weight	25.49	25.34	25.40	23.76			
2020	Standard SHIW weight	17.80	22.91	26.77	32.52			20.25
Baseline redesign adjusted weights								
2020	B1 Cell weight	19.82	23.40	27.99	28.78			15.22
2020	B2 Raking	19.68	23.34	27.83	29.16			15.63
2020	B3 IPW	20.61	23.75	27.94	27.71			12.95
Compound redesign adjusted weights								
2020	C1 Cell weight	24.93	27.50	25.97	21.60			5.45
2020	C2 Raking	24.40	28.43	25.64	21.53			6.66
2020	C3 IPW	23.72	29.11	25.37	21.80			7.54
		Household main income earner: occupation						
Year	Weight	Other	Employee					Bias
2016	Standard SHIW weight	53.38	46.62					
2020	Standard SHIW weight	61.67	38.33					16.58
Baseline redesign adjusted weights								
2020	B1 Cell weight	58.13	41.87					9.50
2020	B2 Raking	58.38	41.62					10.01
2020	B3 IPW	57.85	42.15					8.94
Compound redesign adjusted weights								
2020	C1 Cell weight	55.65	44.35					4.54
2020	C2 Raking	53.10	46.90					0.56
2020	C3 IPW	53.59	46.41					0.43

Share of income earners within the household				
Year	Weight	Less than 50 per cent	More than 50 per cent	Bias
2016	Standard SHIW weight	15.04	84.96	
2020	Standard SHIW weight	12.44	87.56	5.20
Baseline redesign adjusted weights				
2020	B1 Cell weight	13.12	86.88	3.84
2020	B2 Raking	13.13	86.87	3.82
2020	B3 IPW	13.49	86.51	3.10
Compound redesign adjusted weights				
2020	C1 Cell weight	13.96	86.04	2.18
2020	C2 Raking	13.33	86.67	3.42
2020	C3 IPW	13.07	86.93	3.95

Financial literacy					
Year	Weight	0	1	2	Bias
2016	Standard SHIW weight	22.64	49.63	27.73	
2020	Standard SHIW weight	19.84	50.43	29.74	5.60
Baseline redesign adjusted weights					
2020	B1 Cell weight	19.96	51.85	28.19	5.35
2020	B2 Raking weight	19.95	51.68	28.37	5.38
2020	B3 IPW	20.16	52.34	27.50	5.42
Compound redesign adjusted weights					
2020	C1 Cell weight plus	21.90	50.96	27.13	2.66
2020	C2 Raking weight plus	21.36	51.14	27.50	3.02
2020	C3 IPW plus	23.03	51.62	25.35	4.75

Source: Authors' elaborations on SHIW 2016 and 2020 data.

Using the proposed weighting adjustments, we analyse the distributions of core variables of the SHIW, i.e. income and wealth, which are the dimensions most affected by the survey redesign. Of course, these phenomena are likely to have changed from 2016 to 2020, especially in light of the crisis due to the Covid-19 pandemic. Adjusting the weights for the redesign aims at capturing the economic variation of income and wealth net of the redesign effect. We analyse the mean and the Gini index of total household income and wealth across the weighting strategies. The same exercise is repeated in terms of equivalent income and wealth to limit the effects of changes in households' composition.[14]

Table 4 shows that in absence of the redesign adjustment the average total income and the Gini index would have increased respectively by 28 per cent and 7 percentage points. The redesign adjusted weights show instead much more limited variations: the average income increases by 4 to 8 per cent according to the baseline weights and by 3 to 5 per cent according to the compound weights, whereas the Gini index remains stable at the 2016 level according to most

[14] Equivalent income (wealth) is equal to the ratio of total household income (wealth) to the number of equivalent adults. The latter is determined using the OECD-modified equivalence scale, which assigns a value of 1 to the household head, a value of 0.5 to each member aged 14 or over, and a value of 0.3 to each member under age 14.

of the adjusted weights. These results are in line with the National Accounts, which show that the average income of households increases by about 4.5 per cent from 2016 to 2020.

Focusing on total wealth, results show that the standard SHIW weights lead to an increase in the mean and in the Gini index respectively equal to 65 per cent and 7 percentage points. Using the baseline adjusted weights, the increase of average wealth ranges from 13 to 20 per cent and the increase of the inequality index from 0 to 1 percentage points. Using the compound adjusted weights, the range for average wealth is from 3 to 5 per cent, and that for the Gini index from 2 to 4 percentage points.

Hence, the compound weights lead to much more limited variations and closer results to those observed in external data than the baseline weights.

Similar conclusions can be derived in terms of both equivalent income and wealth.

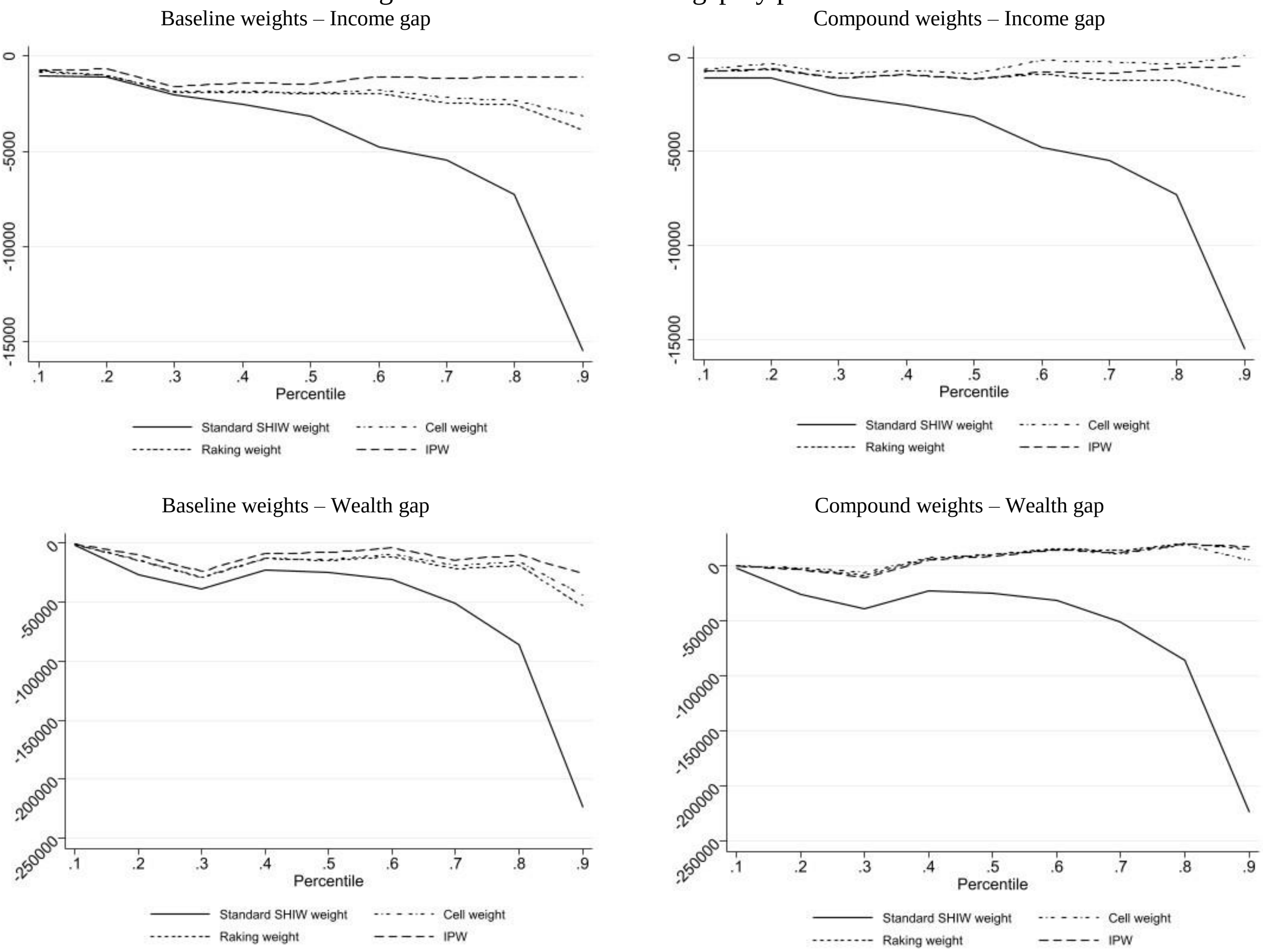
Table 4: income and wealth distributions

			Total income		Total net wealth	
Year	Weight	Obs	Mean	Gini	Mean	Gini
2016	Standard SHIW weight	7,420	30,715	0.358	206,422	0.616
2020	Standard SHIW weight	6,239	39,343	0.428	341,044	0.682
Baseline redesign adjusted weights						
2020	B1 Cell weight	6,239	33,035	0.356	244,215	0.624
2020	B2 Raking	6,239	33,285	0.358	248,669	0.627
2020	B3 IPW	6,173	32,014	0.348	232,568	0.619
Compound redesign adjusted weights						
2020	C1 Cell weight	6,239	31,581	0.354	217,702	0.652
2020	C2 Raking	6,239	32,383	0.358	215,066	0.646
2020	C3 IPW	6,193	31,854	0.353	213,068	0.639
			Equivalent income		**Equivalent net wealth**	
Year	Weight	Obs	Mean	Gini	Mean	Gini
2016	Standard SHIW weight	7,420	18,584	0.335	122,784	0.615
2020	Standard SHIW weight	6,239	24,315	0.395	208,171	0.675
Baseline redesign adjusted weights						
2020	B1 Cell weight	6,239	20,201	0.331	147,110	0.631
2020	B2 Raking weight	6,239	20,345	0.333	149,609	0.633
2020	B3 IPW weight	6,173	19,575	0.323	140,148	0.628
Compound redesign adjusted weights						
2020	C1 Cell weight plus	6,239	19,275	0.330	130,766	0.656
2020	C2 Raking weight plus	6,239	19,741	0.333	128,993	0.651
2020	C3 IPW weight plus	6,193	19,420	0.328	127,460	0.646

Source: Authors' elaborations on SHIW 2016 and 2020 data.

We deepen our analysis by looking at the differences between 2020 and 2016 over the entire income and wealth distributions using the different sets of weights. Figure 1 confirms that the weighting strategies lead to very similar results within the two frameworks (baseline and compound) and that adding variables to the weighting process further decreases the income and wealth gap across the two waves. These results confirm the robustness of the weighting adjustments in providing comparable results along the income and wealth distributions.

Figure 1: income and wealth gap by percentiles



Source: Authors' elaborations on SHIW 2016 and 2020 data. The figure shows the average income and wealth gaps in each percentile of the income and wealth distributions respectively. Income and wealth gaps are obtained as the difference between income and wealth estimated in 2016 (with the standard SHIW weights) and those estimated in 2020 (with the different weighting strategies).

5. Conclusions

To compare the estimates from SHIW 2020 with the previous waves, i.e. net of the sampling redesign, we revise the standard weighting process for the 2020 wave. The weights are adjusted such that the distribution of the 2020 sample with respect to the new design variables (debt and income), conform to the distribution observed in the 2016 sample, net of changes in the population distribution (baseline redesign adjusted weights). We use three different weighting approaches: cell weighting, raking, and inverse probability weighting (IPW). We further enrich this framework by considering, for each of these approaches, not only the new design variables but also other dimensions, correlated to the former, which would have remained stable between 2016 and 2020 in absence of the sampling redesign (compound redesign adjusted weights).

We compare the different strategies in terms of underlying assumptions, efficiency loss, and reduction bias. We also analyse the redesign adjusted weighted income and wealth distributions.

In terms of underlying assumptions and efficiency loss, we conclude that the most suitable approach is raking. Cell weighting leads to a higher weight variability, whereas with the IPW the common support assumption is not always satisfied. Compound adjusted weights, although leading to a slightly higher weight variability, decrease the bias across all the distributions much more than the baseline framework. Focusing on the compound version and redesign strata, raking minimises the bias more than cell weighting and IPW. Turning to the additional variables, cell weighting and IPW seem to minimise the bias depending on the distribution considered, however raking reduction bias shows values very close to the minimum between the other two approaches.

The analysis of income and wealth distributions in 2020 reveals that the redesign adjusted weights correct the variation due to the change in methodology, which would have otherwise soared with the standard SHIW weights. Moreover, estimates of income and wealth distributions are robust across the different weighting strategies.

References

Ballin, M., and G. Barcaroli (2013), “Joint determination of optimal stratification and sample allocation using genetic algorithm”. *Survey Methodology, 39*(2), 369-393

Banca d’Italia (2020), *The Survey on Households Income and Wealth*, Methods and Sources: Methodological Notes.

Barcaroli, G., Ilardi, G., Neri, A., and T. Tuoto (2021), “Optimal sampling design for household finance surveys using administrative income data”. Istituto Nazionale di Statistica, Rivista di statistica ufficiale, 2/2021.

D'Alessio G., and Faiella I. (2002), “Non-response behaviour in the Bank of Italy's Survey of Household Income and Wealth”, Banca d'Italia, Temi di Discussione (Working Papers), 462.

D'Alessio G. and A. Neri (2015), “Income and wealth sample estimates consistent with macro aggregates: some experiments”, Banca d'Italia, Questioni di Economia e Finanza (Occasional Papers), 272.

D'Aurizio L., I. Faiella, S. Iezzi and A. Neri (2006), “The under-reporting of financial wealth in the Survey on Household Income and Wealth”, Banca d'Italia, Temi di Discussione (Working Papers), 610.

Deville, J.C. and C.E. Sarndal (1992), “Calibration Estimators in Survey Sampling.” *Journal of the American Statistical Association* 87, no. 418.

Di Salvatore A., Ilardi G. and A. Neri (2020), “L’uso della Centrale dei rischi per migliorare la qualità delle stime del debito basate sull’Indagine sui bilanci delle famiglie italiane”, mimeo Banca d’Italia.

DiNardo, J., N. M. Fortin and T. Lemieux (1996), “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach”, *Econometrica*, 64(5), 1001-44.

Faiella, I., and R. Gambacorta (2007), “The weighting process in the SHIW”, Banca d'Italia, Temi di Discussione (Working Papers), 636.

Frolich, M. (2007), “Propensity score matching without conditional independence assumption with an application to the gender wage gap in the United Kingdom”, *The Econometrics Journal*, 10, 359–407.

HFCN (2020), *The Household Finance and Consumption Survey: methodological report for the 2017 wave* (No. 35), ECB Statistics Paper.

Iannacchione, V.G., Milne, J.G. and R.E. Folsom (1991), Response probability weight adjustments using logistic regression. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 637-642.

Kalton, G. and I. Flores Cervantes (2003), “Weighting Methods”, *Journal of Official Statistics*, Vol. 19, No. 2, pp. 81-97.

Kalton, G. and D. Kasprzyk (1986), “The treatment of missing survey data”, *Survey Methodology*, Vol. 12, No. 1, pp. 1-16.

Kim, J.K., and J.J. Kim (2007), Nonresponse weighting adjustment using estimated response probability. *Canadian Journal of Statistics*, 35, 501-514.

Kim, J. K., and M.K. Riddles (2012), Some theory for propensity-score-adjustment estimators in survey sampling. *Survey Methodology*, 38(2), 157-165.

Kish, L. (1965), *Survey Sampling*, New York, Wiley, reprinted in 1995.

Kish, L. (1992), "Weighting for Unequal Pi", *Journal of Official Statistics*, Vol. 8, No. 2, pp. 183-200.

Li, F., K. L. Morgan and A. M. Zaslavsky (2018), "Balancing covariates via propensity score weighting", *Journal of the American Statistical Association*, 113, 390–400.

Little, R. J. (1986), "Survey nonresponse adjustments for estimates of means", *International Statistical Review*, 54, 139-157.

Little, R. J. A., Lewitzky, S., Heeringa, S., Lepkowski, J., and R. C. Kessler (1997), "Assessment of weighting methodology for the National Comorbidity Survey", *American journal of epidemiology*, 146(5), 439-449.

Little, R. J., and S. Vartivarian (2005), "Does weighting for nonresponse increase the variance of survey means?", *Survey Methodology*, 31(2), 161.

Neri, A. and M. G. Ranalli (2011), "To misreport or not to report? The measurement of household financial wealth", *Statistics in Transition*, 12-2:281–300.Neri A. and R. Zizza (2010), "Income reporting behaviour in sample surveys", Banca d'Italia, Temi di Discussione (Working Papers), 777.

Rosenbaum, P. and D. Rubin (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika* 70(1): 41–50.

Rothbaum, J. (2017), "Bridging a Survey Redesign Using Multiple Imputation: An Application to the 2014 CPS ASEC", *Journal of Official Statistics* (JOS), 33(1).

Rubin, D. (1974), "Estimating causal effects to treatments in randomised and nonrandomised studies", *Journal of Educational Psychology* 66: 688–70.

Sànchez Muñoz, C. (2011), The Euro-area Household Finance and Consumption Survey: Survey Mode, Oversampling Wealthy Households and Other Methods to Reduce Non-Response Bias. In *UNECE Conference of European Statisticians*.

van den Brakel, J., X. Zhang, and S. M. Tam (2020), "Measuring discontinuities in time series obtained with repeated sample surveys", *International Statistical Review*, 88(1), 155-175.

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

Appendix

A.1 Inverse probability weighting

Using the standard potential outcome framework of the literature on the causal treatment effect (Rubin 1974), denote with N the total population among which each individual i ($i = 1, \dots, N$) is assigned to the treatment, $T_i = 1$, or the control group $T_i = 0$. The potential outcomes are defined as $Y_i(T_i)$. Since only one of the potential outcomes is observed for each i, it is possible to estimate the average treatment effect:

$$\tau_{ATE} = E[Y(1) - Y(0)]$$

In non-experimental studies, it is possible to consistently estimate τ_{ATE} using the propensity score matching method proposed by Rosenbaum and Rubin (1983), which relies on two main assumptions:

- unconfoundedness: conditional on a given observable set of covariates, X, potential outcomes are independent of the treatment assignment:

$$Y(1), Y(0) \perp T |\, X$$

To deal with the potential high dimensionality of X, Rosenbaum and Rubin (1983) propose using the propensity score, i.e. the probability of i to receive the treatment $P(T = 1|X) = e(X) = E[T = 1|X]$. They show that the unconfoundedness is satisfied also conditional on the propensity score:

$$Y(1), Y(0) \perp T |\, P(X)$$

- overlap: the probability of receiving the treatment is bounded away from zero and one, ensuring that individuals with the same values of the covariates have a positive probability of being both in the treatment and in the control group:

$$0 < P(T = 1|X) < 1$$

One way to consistently estimate τ_{ATE} with the propensity score is to weight treated and control individuals to make them representative of the population of interest. In the case of the τ_{ATE}, the target population is the combined treatment and control groups and it can be shown that:

$$\tau_{ATE} = E\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right]$$

It follows from:

$$E\left[\frac{TY}{e(X)}\right] = E\left[\frac{TY}{e(X)}|X\right] = E\left[\frac{1}{e(X)}E[TY|X]\right] = E\left[\frac{E[T=1|X]}{e(X)}E[Y(1)|X]\right] = E[E[Y(1)|X]]$$
$$= E[Y(1)]$$

The IPW estimator of the average treatment effect is:

$$\hat{\tau}_{IPW,ATE} = \frac{1}{N}\left\{\sum_{i=1}^{N}\frac{T_iY_i}{e(X_i)} - \sum_{i=1}^{N}\frac{(1-T_i)Y_i}{1-e(X_i)}\right\} = \frac{1}{N}\left\{\sum_{i=1}^{N}T_iY_iw_1(X_i) - \sum_{i=1}^{N}(1-T_i)Y_iw_0(X_i)\right\}$$

Where the inverse probability weights, which correspond to the Horwitz-Thompson weights, are:

$$\begin{cases} w_1(X_i) = \dfrac{1}{e(X_i)} \; if \; T_i = 1 \\ w_0(X_i) = \dfrac{1}{1-e(X_i)} \; if \; T_i = 0 \end{cases}$$

Li et al. (2018) generalise this approach by proposing a general class of weights, the balancing weights, which incorporate the propensity score and weight of the treatment and control groups to different target populations of interest depending on the estimand, on top of the combined treatment and control groups as in the case of the average treatment effect.

In our context, the target population is the sample of SHIW 2016 and the treatment is being interviewed in 2016, whereas the sample of SHIW 2020 represents the control group. Our goal is to weight the sample of SHIW 2020 as if it was interviewed in 2016. This corresponds to extrapolating the weights under the estimand of the average treatment effect on the treated (ATT), where the target population is the treated group. The ATT is defined as:

$$\tau_{ATT} = E[Y(1)|T=1] - E[Y(0)|T=1]$$

Where the second term represents the average outcome of the control group if it received the treatment. In this case, the IPW estimator of the ATT is:

$$\hat{\tau}_{IPW,ATT} = \frac{1}{N}\left\{\sum_{i=1}^{N}T_iY_i - \sum_{i=1}^{N}\frac{(1-T_i)Y_ie(X_i)}{1-e(X_i)}\right\} = \frac{1}{N}\left\{\sum_{i=1}^{N}T_iY_iw_1(X_i) - \sum_{i=1}^{N}(1-T_i)Y_iw_0(X_i)\right\}$$

Where the inverse probability weights are:

$$\begin{cases} w_1(X_i) = 1 \; if \; T_i = 1 \\ w_0(X_i) = \dfrac{e(X_i)}{1-e(X_i)} \; if \; T_i = 0 \end{cases}$$

A.2 Compound redesign adjusted weights

In the following tables we compare, for each of the weighting strategies, the compound weights in terms of efficiency loss and reduction bias using three different sets of additional variables: set 1 includes only quartiles of property income and the share of employed workers; in set 2 we add the share of income earners within the household, categorized in two classes; set 3, which is the full specification, includes also the financial literacy indicator in three classes.

Table A.1 shows that increasing the number of additional variables does not lead to a higher efficiency loss, except for cell weighting. Raking and IPW shows stable values of the efficiency loss as additional constraints are added.

Table A 1: weight distribution and efficiency loss

Year	Weight	Obs	Mean	SD	Min	Max	L_w	ΔL_w
Compound redesign adjusted weights SET1								
2020	Cell weight	6,239	1	1.39	0.00	7.64	1.93	0.29
2020	Raking	6,239	1	1.38	0.00	7.77	1.89	0.27
2020	IPW	6,190	1	1.39	0.00	7.95	1.93	0.29
Compound redesign adjusted weights SET2								
2020	Cell weight	6,239	1	1.44	0.00	7.83	2.06	0.38
2020	Raking weight	6,239	1	1.38	0.00	7.75	1.90	0.28
2020	IPW weight	6,190	1	1.39	0.00	7.90	1.92	0.29
Compound redesign adjusted weights SET3								
2020	Cell weight	6,239	1	1.47	0.00	8.07	2.15	0.44
2020	Raking weight	6,239	1	1.38	0.00	7.68	1.90	0.27
2020	IPW weight	6,193	1	1.38	0.00	7.96	1.91	0.28

Source: Authors' elaborations on SHIW 2016 and 2020 data.

Table A.2 shows that the different sets lead to very similar results. However, set 2 and set 3 lead to a higher reduction bias than set 1 in a higher number of the analysed dimensions. In particular, the bias of income strata, which is the most relevant redesign variable, is minimised by set 3 with raking. In the other cases, the set that minimises the bias differs according to the weighting strategy. However, set 3 reduction bias is closer to the set that minimises the bias than the other set maximising in most of the dimensions and according to all the three weighting strategies.

In light of these results, our preferred specification is given by set 3.

Table A 2: weighted distributions and reduction bias

Redesign variables							
	Income strata						
	1	2	3	4	5		Bias
Compound redesign adjusted weights SET1							
2020 Cell weight	86.68	8.56	3.50	0.30	0.96		2.94
2020 Raking weight	86.33	8.95	3.42	0.30	0.99		2.25
2020 IPW	87.76	8.25	2.86	0.30	0.83		5.10
Compound redesign adjusted weights SET2							
2020 Cell weight	86.89	8.65	3.30	0.26	0.90		3.36
2020 Raking weight	86.21	9.17	3.34	0.30	0.98		2.00
2020 IPW	87.75	8.22	2.90	0.30	0.83		5.08
Compound redesign adjusted weights SET3							
2020 Cell weight	87.39	8.48	2.91	0.30	0.92		4.36
2020 Raking weight	86.14	9.21	3.36	0.31	0.99		1.87
2020 IPW	87.62	8.32	2.91	0.31	0.84		4.82
	Debt strata						
	0	1	2	3	4	5	Bias
Compound redesign adjusted weights SET1							
2020 Cell weight	79.96	2.64	2.79	2.99	3.71	7.90	4.83
2020 Raking weight	80.01	2.39	2.88	3.11	3.80	7.81	4.91
2020 IPW	82.51	2.33	2.77	2.80	3.48	6.10	3.14
Compound redesign adjusted weights SET2							
2020 Cell weight	80.62	2.49	2.97	2.90	3.75	7.27	3.64
2020 Raking weight	80.56	2.39	2.82	3.06	3.70	7.48	3.95
2020 IPW	82.25	2.35	2.81	2.82	3.53	6.25	3.00
Compound redesign adjusted weights SET3							
2020 Cell weight	81.70	2.24	2.60	2.76	3.78	6.92	3.76
2020 Raking weight	80.53	2.39	2.83	3.06	3.69	7.51	4.00
2020 IPW	82.16	2.39	2.84	2.82	3.53	6.26	2.85
Additional variables							
	Property income class						
	1	2	3	4			Bias
Compound redesign adjusted weights SET1							
2020 Cell weight	24.44	27.96	25.79	21.81			6.01
2020 Raking weight	24.54	28.54	25.50	21.42			6.59
2020 IPW	23.88	29.06	25.39	21.67			7.44
Compound redesign adjusted weights SET2							
2020 Cell weight	24.54	28.05	26.10	21.32			6.80

2020	Raking weight	24.44	28.45	25.64	21.46	6.70
2020	IPW	23.85	29.10	25.35	21.70	7.52
Compound redesign adjusted weights SET3						
2020	Cell weight	24.93	27.50	25.97	21.60	5.45
2020	Raking weight	24.40	28.43	25.64	21.53	6.66
2020	IPW	23.72	29.11	25.37	21.80	7.54
		Household main income earner: occupation				
		Other	Employee			Bias
Compound redesign adjusted weights SET1						
2020	Cell weight	56.32	43.68			5.89
2020	Raking weight	53.44	46.56			0.13
2020	IPW	53.98	46.02			1.21
Compound redesign adjusted weights SET2						
2020	Cell weight	55.97	44.03			5.19
2020	Raking weight	53.09	46.91			0.57
2020	IPW	53.64	46.36			0.53
Compound redesign adjusted weights SET3						
2020	Cell weight	55.65	44.35			4.54
2020	Raking weight	53.10	46.90			0.56
2020	IPW	53.59	46.41			0.43
		Share of income earners within the household				
		Less than 50 per cent	More than 50 per cent			Bias
Compound redesign adjusted weights SET1						
2020	Cell weight	13.39	86.61			3.31
2020	Raking weight	12.67	87.33			4.75
2020	IPW	13.58	86.42			2.93
Compound redesign adjusted weights SET2						
2020	Cell weight	14.00	86.00			2.09
2020	Raking weight	13.38	86.62			3.33
2020	IPW	13.09	86.91			3.91
Compound redesign adjusted weights SET3						
2020	Cell weight	13.96	86.04			2.18
2020	Raking weight	13.33	86.67			3.42
2020	IPW	13.07	86.93			3.95
		Financial literacy				
		0	1	2		Bias
Compound redesign adjusted weights SET1						
2020	Cell weight	20.82	52.24	26.94		5.21
2020	Raking weight	20.96	52.09	26.95		4.91
2020	IPW	21.03	52.34	26.63		5.41

Compound redesign adjusted weights SET2					
2020	Cell weight	20.86	52.38	26.75	5.50
2020	Raking weight	20.87	52.16	26.97	5.05
2020	IPW	21.03	52.29	26.68	5.31
Compound redesign adjusted weights SET3					
2020	Cell weight	21.90	50.96	27.13	2.66
2020	Raking weight	21.36	51.14	27.50	3.02
2020	IPW	23.03	51.62	25.35	4.75

Source: Authors' elaborations on SHIW 2016 and 2020 data.