STATISTICS IN TRANSITION-new series, October 2011
Vol. 12, No. 2, pp. 281—300

TO MISREPORT OR NOT TO REPORT? THE CASE OF THE ITALIAN SURVEY ON HOUSEHOLD INCOME AND WEALTH

Andrea Neri[1], **M. Giovanna Ranalli**[2]

ABSTRACT

The objective of the paper is to adjust for the bias due to unit nonresponse and measurement error in survey estimates of total household financial wealth. Sample surveys are a useful source of information on household wealth. Yet, survey estimates are affected by nonsampling errors. In particular, when it comes to household wealth, unit nonresponse and measurement error can severely bias the estimates. Using the Italian Survey on Household Income and Wealth, we exploit the available auxiliary information in order to assess the magnitude of such a bias. We find evidence that for this kind of surveys, nonsampling errors are a major issue to deal with, possibly more serious than sampling errors. Moreover, in the case of SHIW the potential bias due to measurement error seems to outweigh by far that induced by nonresponse.

Key words: Unit Nonresponse; Measurement Error; Auxiliary Information; Subsampling; Imputation.

1. Introduction

Information on household financial wealth plays an important role in policy analysis. The information available from National Financial Accounts (NFAs) does not usually fulfill policy makers needs since it does not allow analysts to take into account household heterogeneity. Sample surveys are generally used to fill such a gap, since they make it possible to evaluate the impact of shocks, policies and institutional changes on various groups of individuals (European Central Bank, 2009). Yet, the measurement of household financial wealth through sample surveys is a difficult task.

1 Economic and Financial Statistics Department, Banca d'Italia, Rome, Italy. E-mail: andrea.neri@bancaditalia.it

2 Department of Economics, Finance and Statistics, University of Perugia, Italy. E-mail: giovanna@stat.unipg.it

The data we use in this work are from the Survey on Household Income and Wealth (SHIW) conducted by the *Banca d'Italia* (the Italian central bank) every two years. The main objective of the SHIW is to study economic behaviors of Italian households. The survey is used both for research and for the evaluation of economic policies. Previous studies show that survey estimates usually underestimate the corresponding aggregate figures. Even if national accounts can hardly be considered flawless, the comparison is useful to highlight some quality issues in the microdata. In general, the main sources of error for this kind of surveys are the low propensity of wealthy households to participate in the survey (D'Alessio and Faiella, 2002) and the measurement error that likely arises when collecting survey data of this type (Biancotti et al., 2008). These issues are particularly relevant when it comes to financial wealth. First, financial assets and liabilities are highly concentrated in the hands of wealthy households. Second, the increasing complexity of household financial portfolios increases the respondents' difficulty in retrieving a correct information.

From a data producer point of view, it is crucial to study all the potential survey error components in order to allocate the limited financial resources where most needed (Biemer, 2010). The objective of the paper is to quantify the magnitude of the two main sources of error (nonresponse and measurement error) on the estimator of total household wealth components using the auxiliary information available for the SHIW survey.

The analysis is based on two steps. We first deal with unit nonresponse. Nonresponse is considered as a second phase of sampling with unknown probabilities (see e.g. Särndal, Swensson and Wretman, 1992, Chapt. 9). To this end, we use individual response propensities estimated using data coming from a survey conducted on a subsample of unwilling-to-participate households and from past surveys for panel households (see Little, 1986; Ekholm and Laaksonen, 1991; Kim and Kim, 2007, when estimation is conducted using logistic models). Secondly, we deal with measurement error using a validation sample made of a survey of customers of a major Italian commercial bank, with survey data matched to the bank's administrative records. Measurement error is considered as a third source of uncertainty modeled using propensities to misreport on the validation sample. These propensities are then used to develop a simulation-based adjustment process for SHIW assets data.

The paper is organized as follows. In Section 2 a description of the sampling design employed for SHIW is provided. Then, Section 3 describes the proposed methodology to tackle nonresponse. Different models are developed and considered to estimate response probabilities for panel and non-panel households according to the available auxiliary information. It will be shown that nonresponse is driven by different factors for the two types of households. Section 4 provides details on the models used to estimate misreporting propensities and to obtain imputed values for the variables of interest. Finally, in Section 5 a comparison of the alternative estimators obtained using the aforementioned techniques is provided, together with an appraisal of the role of the auxiliary information employed for nonresponse and measurement error adjustments on the estimates for the survey at hand. Some concluding remarks are also provided to envision further and more general methodological developments suggested by the present application.

2. The sampling design used for the SHIW

The SHIW is a two stage survey, with municipalities and households as primary and secondary sampling units, respectively. PSUs are stratified by administrative region (NUTS 1 level) and population size (less than 20,000 thousand inhabitants; between 20,000 and 40,000; 40,000 or more). Within each stratum, PSUs are selected to include all those with a population of 40,000 inhabitants or more and those with panel households (self-representing municipalities), while smaller municipalities are selected using probability proportional to size sampling (without replacement). Individual households are then randomly selected from administrative registers.

Up to 1987 the survey was conducted with time-independent samples (cross sections) of households. In order to make it possible to analyze the change in the phenomena under investigation, since 1989 part of the sample has included households interviewed in previous surveys (panel households). The overall sample size for the 2008 edition is 7,977 households, with 4,345 panel households (54.5% of the sample). The rotation scheme for the panel component is as follows: households that had participated for at least two waves are all included in the sample, while the remaining panel households are selected randomly from among those interviewed only in the previous survey. As a result, the longitudinal component of the sample consists of a quite heterogeneous group of households as of the year of the first interview and the number of waves. For example, of the 4,345 panel households in 2008, 28 have participated since 1987, 146 since 1989, 347 since 1991 and 1,143 come from the previous 2006 edition.

The questionnaire used in the survey has a modular structure. It is made of a general part addressing aspects concerning all households and a series of additional sections containing questions that are relevant to specific subsets of households. Data collection is entrusted to a specialized company using about 200 professional interviewers. Substitutions are allowed under a strict protocol. In particular, interviewers have no influence on when a household can be dropped and which household to use as a substitute. Information is collected using the Computer-Assisted Personal Interviewing (CAPI) technique. Interviews last an average of 55 minutes. In addition, interviews are considered valid if they have no missing items on the questions regarding income and wealth. As a result, item nonresponse is negligible while, as it will become clearer soon, unit nonresponse is a major issue.

3. Unit nonresponse

In 2008, 14,209 households have been contacted and 7,977 have been interviewed (56.1%), while 32.4% has refused to cooperate and the remaining 11.5% is given by non-contacts (see Table 1). Nonresponse affects particularly non-panel households, in fact 41.1% refuses to participate in the survey, while this percentage decreases for panel households to 18.5%. To study the factors that drive nonresponse and try to adjust for nonresponse bias, we use a two-phase approach: the selected sample is considered as the first phase sample, while the set of respondents is considered as a second phase sample. Each unit in the population has attached a probability of inclusion for such second phase sample, that is a response probability and, therefore, an unknown characteristic.

Table 1. Households contacted in 2008 and reasons for non-participation.

	Panel		Non-panel		Total	
	Number	%	Number	%	Number	%
Respondents	4,345	79.3	3,632	41.6	7,977	56.1
Refusals	1,012	18.5	3,589	41.1	4,601	32.4
Not at home	120	2.2	1,511	17.3	1,631	11.5
Total	5,477	100.0	8,732	100.0	14,209	100.0
Ineligible*	150	2.7	629	6.7	779	5.2

* Households not found at their address (wrong address, death, change of address).

More formally, given a finite population of N elements $U = \{1,\ldots,k,\ldots,N\}$, the aim is to estimate the vector of totals $\mathbf{t}_y = \sum_U \mathbf{y}_k$, where $\mathbf{y}_k$ is the value of the p-dimensional vector of variables of interest $\boldsymbol{y}$ for the k-th unit. We will use in general the shorthand $\sum_A$ for $\sum_{k\in A}$, with $A \subseteq U$ an arbitrary set. In our application $p = 6$: for each of two types of aggregated financial assets and for financial liabilities we have the number of households possessing the asset (or the liability) and the amount possessed. The two types of aggregated assets are: bonds (government + private bonds) and risky assets (shares + mutual funds + managed savings).

A sample s of size n is drawn from U according to the sampling design $p(s)$ that induces first order inclusion probabilities $\pi_k = P(k \in s)$. Since nonresponse occurs, the response set r of size n_r is obtained from the response mechanism given by the distribution $q(r \mid s)$, with $r \subseteq s$ and $n_r \le n$. Let $\delta_k = 1$ if unit k responds and zero otherwise. Then, $\theta_k = P(k \in r \mid k \in s) = P(\delta_k = 1)$ is the probability that unit k responds given that it was included in the sample. Since θ_k is considered as an individual characteristic defined for all units in the populations, $\theta_k = P(k \in r \mid k \in s) = P(k \in r)$. If these probabilities were known, the two-phase estimator

$$\hat{\mathbf{t}}_{y,2} = \sum_r \frac{\mathbf{y}_k}{\pi_k \theta_k}$$

would be unbiased for $\mathbf{t}_y$.

When auxiliary information is available for all units in s, these probabilities can be estimated using response propensties. One of the most common and simple technique to handle nonresponse is given by constructing response homogeneity groups: the population (or the sample s) is partitioned into groups such that units belonging to the same group are assumed to have the same response propensity. In the SHIW such propensities are currently estimated for a PSU l by the ratio between the effective number of components in the respondents set m_{lr} and the number of components in the original sample m_{ls}. Therefore, the estimated response propensity for household k is given by $\hat{\theta}_k^S = m_{l(k)r}/m_{l(k)s}$, with $l(k)$ denoting the PSU to which household k belongs to. Then, the estimator of the total is computed as

$$\hat{\mathbf{t}}_{y,\ SHIW} = \sum_r \frac{\mathbf{y}_k}{\pi_k \hat{\theta}_k^S}. \tag{1}$$

Another common but more flexible approach is to use a logistic model for the response indicator δ_k under the assumption of the classical binomial response model that δ_k is independent of δ_j for $k \neq j$, i.e. $\theta_{kj} = P(k \;\&\; j \in r) = \theta_k \theta_j$ (Little, 1986; Ekholm and Laaksonen, 1991). More in general, the response probability can be assumed to be the inverse of a known *link* function of an unknown (but estimable) linear combination of model variables (Folsom, 1991; Fuller et al., 1994; Kott, 2006). Asymptotic properties in the case of a logistic link are explored in (Kim and Kim, 2007). This is a reasonable approach here too, because the design foresees the sampling of full households. Note that response homogeneity groups and logistic models provide the same response propensities when the auxiliary variables used in the logistic model are the response group indicator variables.

In this application, two different models and data sources have been employed for panel and for non-panel households. In particular, we can partition the original sample s (and the respondents set r) into two sub-samples given by s_p and s_{np} (and by r_p and r_{np}) corresponding to panel and non-panel households, respectively, so that $s_p \cup s_{np} = s$ (and $r_p \cup r_{np} = r$). Once models are selected, estimates of θ_k for $k \in r_p$ and for $k \in r_{np}$ are obtained and denoted by $\hat{\theta}_k^M$. The estimator of the total is then computed as

$$\hat{\mathbf{t}}_{y,\ NR} = \sum_r \frac{\mathbf{y}_k}{\pi_k \hat{\theta}_k^M}. \tag{2}$$

3.1. Response model for panel households

To estimate response probabilities for panel households we exploit the information from the previous interview(s) and use additive logistic regression (Ruppert, Wand and Carrol, 2003). In particular, the effect of the age of the interviewer has been modeled using nonparametric regression via p-splines given that there was evidence of a more complex relationship than a linear one. Figure 1 shows the shape of the effect of the age of the interviewer on the linear predictor scale. In general, younger interviewers tend to obtain lower response rates. Table 2 shows the coefficients for the other variables found significant through model selection from all those available.

Figure 1. The estimated effect of the age of the interviewer (and 95% confidence bounds) on the linear predictor scale from the additive logistic response probability model for panel households.

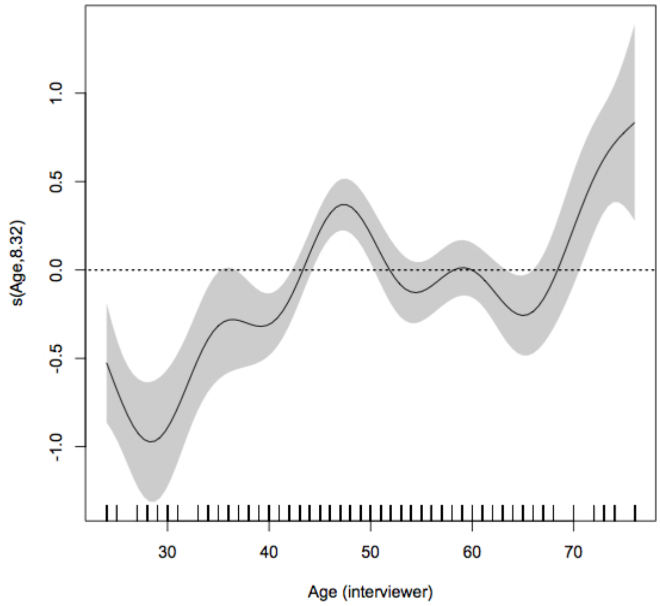

Table 2. Logistic response probability model for panel households - estimated coefficients, standard errors and *p*-values.

Variables	Coeff	Std. Err.	p-value
Intercept	-1.48	0.21	< .001
Municipalities with more than 500,000 inhabitants	-0.58	0.12	< .001
Household living downtown	0.34	0.10	< .001
Number of waves (household)	0.18	0.02	< .001
Number of members of household	0.11	0.03	< .001
High level of education (interviewer)	0.34	0.10	< .001
Number of waves (interviewer)	0.03	0.01	< .001
Good climate at previous interview	0.20	0.02	< .001
Workload of interviewer 21 - 100	-0.06	0.09	0.481
Workload of interviewer 101 - 300	-0.43	0.13	< .001
Workload of interviewer > 300	0.50	0.15	< .001

Pseudo R^2 = 0.085; 5,625 obs.

For panel households, responding appears to be mainly a matter of trust. On the contrary, household economic conditions, although included in the model as available auxiliary information, do not show to have an effect on the response propensity. The only household's attributes that have an effect on the response rate, are the number of members and the place where they live: numerous households and those living downtown are more likely to continue to participate, while those living in larger municipalities show higher attrition. On the other side, a major determinant of response propensity is the number of waves the household has already been interviewed successfully. Old panel households are more willing to continue to participate. For instance, households who have entered the panel in 2006 have an estimated response probability of about 0.68. This figure jumps to about 0.90 for those households who have been in the panel for more than 5 waves. One reason is the building of a relationship of trust between respondents and the survey and, in particular, with the interviewer. Households become progressively aware that there is no risk of a break in confidentiality. At the same time, their identification with survey aims increases as time passes. In order to preserve such a link with the respondents, panel households are usually assigned to the same interviewer.

Moreover, a climate judged as "good" by the interviewer at the previous interview provides higher household cooperation. Other important variables affecting response are connected to the characteristics of the interviewer. Interviewers with a relatively higher degree of education, who take larger workloads and have participated in a larger number of editions of the survey have better results. The estimated function of age and coefficients from Table 2 are used to predict response probabilities $\hat{\theta}_k^M$ for all $k \in r_p$, i.e. for the 4,345 interviewed panel households to be used in estimator (2).

3.2. Response model for non-panel households

In 2008, 8,732 non-panel households have been contacted and 3,632 (41.6%) have been interviewed. About 70% of the 5,100 non participating households has explicitly refused to cooperate, while the remaining 30% is not found at the address. The response propensity modeling for non-panel households is based on an *ad hoc* source of auxiliary information. Starting from 2006, the survey agency has to carry out a survey among non-panel households that refuse to participate. It is a telephone survey (CATI technique) run on a sample of non-respondents. This survey is conducted during the fieldwork, while trying to convert refusals. If the attempt is not successful, the interviewers ask whether the household is at least willing to reply to a five minutes telephone questionnaire. The survey agency has to contact all the non-participating households. Among non-participating households, only 316 have agreed to the telephone interview, about 6% of those households which are selected but do not participate.

For non-panel households, auxiliary information is not known for each unit in the original sample s_{np}, but only for the respondents r_{np} and a subsample of units of $s_{np} \setminus r_{np}$. Nonetheless, we propose to estimate response probabilities using weighted logistic regression on a dataset made of the subsample of nonrespondents and the sample of respondents. In general, (i) nonrespondents should be given a weight equal to the inverse of the inclusion probability coming from the sub-sampling design, while (ii) respondents a weight equal to 1. For the case at hand, given that the sub-sample of nonrespondents is not a probabilistic sample, a sort of post-stratification is employed in which (i) nonrespondents are given a weight that sums up to the total number of nonrespondents by geographical area and size of the municipality resulting from the sample register file, while (ii) respondents are given a weight equal to 1 (Laaksonen and Chambers, 2006, use a similar approach when the variable of interest is observed on a sub-sample of non-respondents – follow-up sample). This approach assumes that sub-sampling is at random and that nonrespondents in the sub-sample can be considered similar to the others in the same post-stratum. We will discuss more in detail later whether such assumptions can be considered valid in the situation at hand.

Response probabilities are then estimated as a function of a set of variables that are available for both samples using additive logistic regression as for panel households. In particular, in this case the effect of the age of the head of the household has been modeled using nonparametric regression via p-splines given that there was evidence of a more complex relationship than a linear one. Figure 2 shows the estimated function of age on the linear predictor scale, while Table 3 shows the estimated coefficients for the other variables found significant.

The propensity to respond decreases steadily with age until the age of 30 where it stabilizes and then decreases again. A slight increase is then detected between 65 and 75. The horizontal dotted line shows that households with heads who are 50 or younger are more willing to participate than those with heads who are older than 50. Table 3 shows that response probabilities decrease for households whose head is self-employed, home owner, graduated, or retired. In addition, households living in the North/Centre of Italy and those with a larger number of members are less willing to participate. On the contrary, response propensity increases for households who in smaller municipalities. Finally, households with two (three or more) wage earners are less (more) likely to respond than those with only one.

Figure 2. The estimated effect of age (and 95% confidence bounds) on the linear predictor scale from the logistic response probability model for non-panel households.

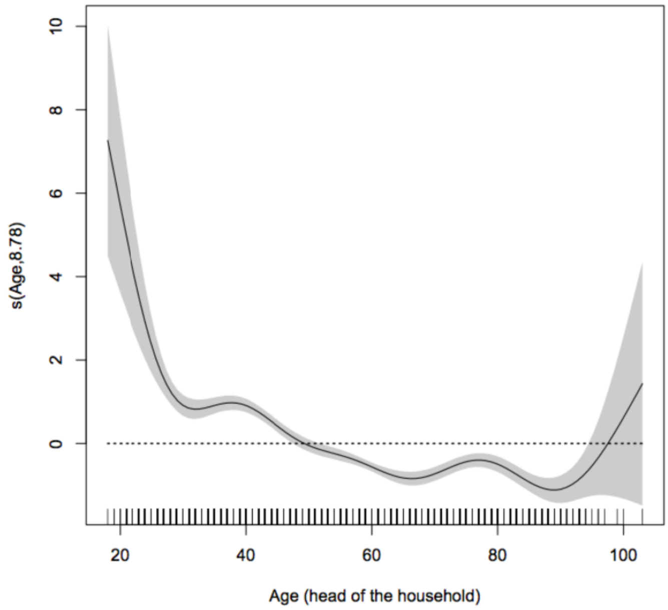

Table 3. Logistic response probability model for non-panel households - estimated coefficients, standard errors and p-values.

Variables *	Coeff	Std. Err.	p-value
Intercept	1.599	0.117	$<$ .0001
Living in the North/Centre of Italy	-0.744	0.058	$<$ .0001
Municipality with less than 500,000 inhabitants	0.362	0.059	$<$ .0001
Household originally selected (vs substitute)	-0.280	0.057	$<$ .0001
Workload of interviewer 21 -- 100	0.361	0.058	$<$ .0001
Workload of interviewer 101 -- 300	0.880	0.077	$<$ .0001
Workload of interviewer $>$ 300	1.912	0.110	$<$ .0001
Self-employed	-0.482	0.084	$<$ .0001
Graduated	-0.265	0.078	0.0007
Retired	-0.291	0.109	0.0073
Home owner	-0.658	0.058	$<$ .0001
Number of members of the household	-0.368	0.028	$<$ .0001
Number of income earners = 2	-0.103	0.055	0.0639
Number of income earners $\geq$ 3	0.525	0.099	$<$ .0001

* Demographic characteristics refer to the head of the household; Pseudo R^2 = 0.357; 3,948 obs.

In the response model for panel households, information about the interviewer was found to be significant. For non-panel households, such information is not available. Indeed the interviewers for the CATI survey are different from those running the CAPI survey. The only useful information is that of the workload of the original interviewer measured by the number of households to be interviewed. Those with a larger workload tend to have larger response rates than the others. A likely explanation for this result is that the survey agencies usually allocate a larger number of households to their best interviewers.

Note that no explicit income related items are surveyed on the sub-sample of nonrespondents given their refusal to participate to the SHIW. Therefore, there is no information available on this to be incorporated in the response model for non-panel household. Nevertheless, some of the variables found significant that are related to the head of the household are usually good predictors of wealthier households (being graduated, self-employed, home owner). The estimated function of age and coefficients are then used to compute estimated response probabilities $\hat{\theta}_k^M$ for all $k \in r_{np}$, i.e. for all 3,632 non-panel respondents to be used in estimator (2).

4. Measurement error

Financial assets collected in the SHIW are also likely to be affected by misreporting of the financial tools and amounts by households. Such misreporting may well be due to a malicious behavior, with underreporting being the most likely outcome. However, it can also be done in *bona fides*, given the respondents' difficulty in retrieving a correct information due to the increased complexity of household financial portfolios. For these reasons, the value for the variables of interest reported by unit k, which we will denote by $\tilde{\mathbf{y}}_k$, may differ from the true value $\mathbf{y}_k$.

Bias caused by measurement error could be adjusted for by selecting a subsample m of the respondents where a more accurate measurement of the study variable(s) is taken (e.g. Lessler and Kalsbeek, 1992). When the subsample is selected using a probabilistic sampling design, the framework is another example of two phase sampling. When nonresponse is present, as is the case of the SHIW, then a three phase framework arises: $m \subset r \subseteq s$ of dimension $n_m < n_r \leq n$ is selected using the design $p_m(m \mid r, s)$ with conditional inclusion probabilities $\tau_k = P(k \in m \mid k \in r)$. Then, the three phase estimator

$$\hat{\mathbf{t}}_{y,3} = \sum_m \frac{\mathbf{y}_k}{\pi_k \theta_k \tau_k}$$

would be unbiased for $\mathbf{t}_y$. Of course the efficiency of $\hat{\mathbf{t}}_{y,3}$ depends on the dimension of m: a compromise choice should be made according on how expensive it is to retrieve the correct information on units. The unbiased estimator $\hat{\mathbf{t}}_{y,3}$ is constructed using the subsample m alone. Other estimators that make a better use of the information on the respondents set r (given by the correlated surrogate variable $\tilde{\mathbf{y}}_k$ and some auxiliary information) can be proposed in a model assisted framework to improve efficiency, using GREG type or model calibration type estimators (e.g. see the hint in Wu and Luan, 2003, Section 6, in a two phase framework). These extensions go however beyond the scope of this paper.

Now, for this survey we have no such data available on a subsample of r and the three phase approach cannot be used as described earlier. However, we have data available from an independent experiment survey carried out by the *Banca d'Italia* and a major Italian bank group on a sample of customers of the latter. The experiment was carried out in 2003 on a sample of 1,681 households where at least one member was a customer of the bank group. In order to get data comparable with that coming from the SHIW, the questionnaire and the survey design were as close as possible to those used in the most recent edition of the SHIW (2002). Interviews were made by the same survey agency using the same interviewers and CAPI technique.

Survey data had then been matched with the bank customers database containing the amount of the assets actually held by the individuals selected in the sample. Since these amounts and those declared in the interview refer to the same period (year 2002), they were fully comparable. The two sets of data are merged through an operation of exact record linkage. The resulting dataset will be referred to as our “validation sample”.

Although temporally misaligned, the validation sample gives us the possibility of studying misreporting behaviors and of trying to extrapolate it to the SHIW sample. This is accomplished in a two-step fashion. In fact, household wealth reporting in surveys is generally a two-stage process involving first the reporting of ownership of assets and liabilities and then the reporting of the amounts owned (Moore et al., 2000). Errors can occur at either of the two stages. An entire financial instrument can be either omitted or reported even if it is not actually owned. Alternatively, the ownership may be reported correctly but the amount can be misreported. Even if the respondent has fully understood the question, he/she may fail to retrieve the correct information. Lack of knowledge is the first cause. Even if in the SHIW the respondent is selected as the more knowledgeable person in the household, he or she may not know the true situation of all the other components.

In the final stage, after recalling the requested information, the respondent adopts a response strategy. Deliberate underreporting is probably the major cause of response error at this stage. Nonetheless, besides deliberate prevarication, there are further possible sources of error, like those coming from the interaction between the interviewer and the respondent. For instance, if the respondent belongs to a very rich household he/she may decide to underreport wealth because of a need of “social conformability” with the interviewer. This could be considered as a special case of the so called "social desirability bias" (Bagozzi, 1994), namely, the tendency for an individual to present himself in a way that makes the person look positive to cultural norms or standards. On the opposite side, over-reporting may arise from a respondent willing to impress the interviewer.

Recall that we consider two types of aggregate financial assets – bonds and risky assets – and also financial liabilities. For each of these three quantities we have, therefore, two related variables: possession and amount possessed. For the former, we have in particular two variables defined as follows:

$$y_{pk} = \begin{cases} 1 & \text{if unit } k \textit{ possesses} \text{ financial instrument } p = 1,2,3 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{y}_{pk} = \begin{cases} 1 & \text{if unit } k \textit{ declares to possesses} \text{ financial instrument } p = 1,2,3 \\ 0 & \text{otherwise} \end{cases}.$$

Table 4. Logistic response probability model for ownership of bonds – estimated coefficients, standard errors and p-values.

Variables *	Coeff	Std. Err.	p-value
Intercept	-4.67	0.98	< .0001
Self-reported ownership of bonds	2.50	0.17	< .0001
Self-reported ownership of risky assets	0.45	0.12	0.0003
Employee	0.21	0.19	0.2589
Self-employed	0.23	0.20	0.2502
Secondary school diploma	0.39	0.15	0.0086
University degree	0.55	0.19	0.0045
Age	0.06	0.03	0.0477
Age^2	-0.01	0.03	0.6520
Living in municipalities with more than 30,000 inhabitants	-0.26	0.13	0.0432
Living in the North/Centre of Italy	0.86	0.20	< .0001
Living in a rural area	-0.51	0.24	0.0315
Living in the town outskirts	-0.36	0.23	0.1133
Number of income earners	0.02	0.05	0.6610
First quartile of household income	-0.05	0.16	0.7792
Fourth quartile of household income	-0.16	0.16	0.3139
First quartile of household real wealth	0.35	0.15	0.0203
Fourth quartile of household real wealth	0.06	0.16	0.7051
Client of more than one bank	0.06	0.13	0.6171
Respondent's level of understanding of the questions	0.01	0.07	0.8899
Respondent's easiness to answer the questions	0.00	0.07	0.9935
Reliability of the information provided by the respondent	-0.03	0.04	0.4168

* Demographic characteristics refer to the head of the household; Pseudo R^2 = 0.364; 1,681 obs.

The validation sample then allows to identify, in the first phase, households among those declaring not to own a given financial asset, that very likely own it and provided incorrect data and, symmetrically, to identify households declaring to own a financial asset, but unlikely to possess it. This is accomplished by estimating a logistic model for $P(y_{pk} = 1)$ using a vector of socio-economic characteristics both at the household and at the head of household level as covariates, together with the declared value $\tilde{y}_{pk}$.

Tables 4, 5 and 6 report the results from the models for the variables bonds, risky assets and financial liabilities, respectively. The tables also show the p-values for the covariates considered and an overall measure of goodness of fit. Note that these models are fit with the aim of imputation. Therefore, model selection is based on the performance for out-of-sample predictions rather than on the amount of variability explained by the covariates. For this reason, also non-significant covariates can be found in the aforementioned Tables. A nonparametric term for the effect of the age of the head of the household has been tested, but provided no better performance than a quadratic function of age and was dropped in all models.

Table 5: Logistic response probability model for ownership of risky assets - estimated coefficients, standard errors and p-values.

Variables *	Coeff	Std. Err.	p-value
Intercept	-2.27	0.89	0.0105
Self-reported ownership of bonds	0.18	0.16	0.2640
Self-reported ownership of risky assets	2.65	0.15	< .0001
Employee	0.56	0.19	0.0038
Self-employed	0.07	0.21	0.7455
Secondary school diploma	0.31	0.15	0.0347
University degree	0.13	0.20	0.5184
Age	0.04	0.03	0.2298
Age^2	-0.01	0.03	0.6452
Municipalities with more than 30,000 inhabitants	0.09	0.13	0.4904
Living in the North/Centre of Italy	0.19	0.17	0.2667
Living in a rural area	0.41	0.25	0.0906
Living in the town outskirts	0.50	0.23	0.0332
Number of income earners	-0.12	0.06	0.0322
First quartile of household income	-0.20	0.16	0.2149
Fourth quartile of household income	0.23	0.17	0.1795
First quartile of household real wealth	-0.01	0.15	0.9243
Fourth quartile of household real wealth	0.45	0.17	0.0092
Client of more than one bank	0.04	0.13	0.7346
Respondent's level of understanding of the questions	-0.07	0.07	0.3640
Respondent's easiness to answer the questions	0.03	0.08	0.6768
Reliability of the information provided by the respondent	-0.04	0.04	0.2953

* Demographic characteristics refer to the head of the household; Pseudo $R^2 = 0.393$; 1,681 obs.

From Table 4, the probability of holding bonds increases for relatively less wealthy households living in the North/Center of Italy and in smaller municipalities. In addition it increases with the age of the head of the household and with his/her level of educational attainment. From Table 5, on the other hand, the probability of owning risky assets increases for relatively wealthier households with few components, which live in residential or rural areas. Finally, Table 6 shows that the probability of owing financial liabilities decreases for relatively less wealthy households living in the North/Centre of Italy, in smaller municipalities and in rural or residential areas.

Table 6. Logistic response probability model for ownership of financial liabilities - estimated coefficients, standard errors and p-values.

Variables *	Coeff	Std. Err.	p-value
Intercept	-5.93	1.94	0.0022
Self-reported ownership of liabilities	4.43	0.30	<.0001
Employee	-0.87	0.39	0.0255
Self-employed	0.40	0.40	0.312
Secondary school diploma	-0.18	0.30	0.5413
University degree	0.05	0.42	0.9063
Age	0.25	0.07	0.0009
Age^2	0.00	0.00	0.0003
Municipalities with more than 30,000 inhabitants	0.10	0.27	0.7211
Living in the North/Centre of Italy	0.00	0.34	0.9894
Living in a rural area	-1.43	0.46	0.0021
Living in the town outskirts	-0.94	0.41	0.0227
Number of income earners	0.12	0.11	0.2999
First quartile of household income	0.15	0.34	0.6562
Fourth quartile of household income	0.41	0.33	0.2146
First quartile of household real wealth	-0.18	0.31	0.5698
Fourth quartile of household real wealth	-0.04	0.35	0.8968
First quartile of household financial assets	0.44	0.29	0.1328
Fourth quartile of household financial assets	-0.73	0.36	0.046
Respondent's level of understanding of the questions	-0.33	0.17	0.0494
Respondent's easiness to answer the questions	0.31	0.17	0.0714
Reliability of the information provided by the respondent	-0.08	0.09	0.3653

* Demographic characteristics refer to the head of the household; Pseudo R^2= 0.403; 949 obs.

In a second phase misreporting on the amount held is estimated through a separate model for each of the three financial tools. In particular, for the three last variables of interest y_p, $p = 4,5,6$, let $r_{pk} = y_{pk}/\tilde{y}_{pk}$ be the ratio of the actual and the declared amount held by household k. Then $\log r_{pk}$ is modeled for each unit of the validation sample on a set of household characteristics, including household income and wealth classes, a synthetic judgmental variable on the reliability of the information provided in the interview expressed by the interviewer (a discrete score ranging from 1 to 10), and the declared amount. Tables 7 and 8 report the results from the models for bonds and for risky assets, respectively. For financial liabilities the number of available observations is too small for proper modeling of r_{pk}. Therefore, a common mean model for the ratio is estimated, that takes value 1.064 for all units in the sample.

These two sets of models can be used to adjust measurement error in the SHIW as follows. If we assume that the misreporting behavior of the households in the bank experiment is the same as that of those in the SHIW, then parameter estimates from these two sets of models can be used to stochastically impute micro data for households in the SHIW (imputation for measurement error correction for distribution function estimation is explored in Durrant and Skinner, 2006). In particular, profiles of households given by unique combinations of covariate values are constructed from the SHIW, then predictions $\hat{\mathbf{y}}_k$ are obtained using parameter estimates from the aforementioned models that substitute the surveyed values $\tilde{\mathbf{y}}_k$. A random error term is then added to preserve variability. In particular, in the models for asset ownership a Bernoulli experiment is conducted to assign the imputed possession of a given asset class. As of the models related to the amount possessed, a random draw from a zero-mean normal distribution is added to the imputed value; the variance of the normal distribution is given by that of the residuals of the model fitted in the validation sample.

Table 7. Regression model for log of ratio between actual and declared amount of *bonds* - estimated coefficients, standard errors and p-values.

Variables *	Coeff	Std. Err.	p-value
Intercept	0.85	0.69	0.2197
Second quartile of household financial wealth in risky assets	-0.63	0.12	<.0001
Third quartile of household financial wealth in risky assets	-0.64	0.12	<.0001
Fourth quartile of household financial wealth in risky assets	-1.09	0.12	<.0001
Employee	0.38	0.13	0.0038
Self-employed	0.25	0.13	0.062
Secondary school diploma	-0.02	0.10	0.8718
University degree	-0.06	0.12	0.6204
Age	-0.02	0.02	0.5034
Age^2	0.00	0.00	0.2413
Living in the North/Centre of Italy	0.23	0.17	0.1666
First quartile of household income	-0.02	0.11	0.8534
Fourth quartile of household income	0.11	0.10	0.2782
First quartile of household real wealth	-0.08	0.11	0.4769
Fourth quartile of household real wealth	0.07	0.10	0.5081
Client of more than one bank	0.03	0.08	0.7526
Reliability of the information provided by the respondent	0.01	0.02	0.703

* Demographic characteristics refer to the head of the household; Pseudo R^2 = 0.453; 482 obs.

Table 8. Regression model for the log of ratio between actual and declared amount of *risky assets*- estimated coefficients, standard errors and p-values

Variables*	Coeff	Std. Err.	p-value
Intercept	0.43	0.31	0.1634
Second quartile of household financial wealth in risky assets	-0.04	0.07	0.5864
Third quartile of household financial wealth in risky assets	-0.34	0.06	<.0001
Fourth quartile of household financial wealth in risky assets	-0.32	0.07	<.0001
Employee	-0.08	0.08	0.3025
Self-employed	-0.04	0.08	0.6245
Secondary school diploma	0.14	0.06	0.0174
University degree	0.16	0.07	0.0318
Age	0.00	0.01	0.9866
Age squared	0.00	0.00	0.802
Living in the North/Centre of Italy	0.00	0.08	0.9887
First quartile of household income	0.13	0.07	0.0578
Fourth quartile of household income	0.12	0.06	0.0416
First quartile of household real wealth	-0.07	0.06	0.2726
Fourth quartile of household real wealth	-0.04	0.06	0.5137
Client of more than one bank	0.00	0.05	0.9368
Reliability of the information provided by the respondent	-0.04	0.01	0.0009

* Demographic characteristics refer to the head of the household; Pseudo $R^2 = 0.126$; 876 obs.

The final estimator of the total of the variables of interest adjusted for measurement error is essentially a two-phase estimator, and takes the two following forms according to whether nonresponse is adjusted for using the logistic models or not:

$$\hat{\mathbf{t}}_{\hat{y},ME} = \sum_r \frac{\hat{\mathbf{y}}_k}{\pi_k \hat{\theta}_k^S}, \tag{3}$$

$$\hat{\mathbf{t}}_{\hat{y},NRME} = \sum_r \frac{\hat{\mathbf{y}}_k}{\pi_k \hat{\theta}_k^M}. \tag{4}$$

5. Results and concluding remarks

In this section we report the final estimates of the six variables of interest (holding and amount possessed for bonds, risky assets and financial liabilities) obtained using the alternative estimators discussed in the previous sections. Table 9 reports the estimates of the total of the first three variables of interest

($p = 1,2,3$), i.e. the number of households holding the financial instrument, plus the estimate for the number of households holding either bonds or risky assets, or holding both (*total financial assets*). Note that the first two estimators are computed as in equations (1) and (2), respectively, in which $\mathbf{y}_k$ is replaced by the observed surrogate value $\tilde{\mathbf{y}}_k$. Table 10 reports, on the other hand, the estimates of the total of the last three variables ($p = 4,5,6$), i.e. the amount held (in billions of Euros) by households for the three financial instruments, plus the estimate of the total financial assets own. As a measure of the coverage of each estimate, the ratio of its value with respect to the corresponding estimate coming from the National Financial Accounts (NFAs) is also computed.

Our main findings may be summarized as follows. Underreporting and unit nonresponse emerge as particularly serious issues with regard to financial assets. The estimator $\hat{\mathbf{t}}_{\hat{y},\,NRME}$ of the total adjusted for both nonresponse and measurement error is about 3-4 times higher that the unadjusted $\hat{\mathbf{t}}_{\tilde{y},\,SHIW}$ when financial assets are considered (Table 9). When it comes to the estimation of the amounts held (Table 10), the bias increases: the SHIW estimates for the variables bonds and risky assets should be inflated by factors from 5 to 8 in order to get the figures obtained with $\hat{\mathbf{t}}_{\hat{y},\,NRME}$. The latter are those closest to the estimates coming from NFAs (last two columns of Table 10).

Correction for nonresponse and measurement error for financial liabilities seems to be less effective. As far as the amount is concerned, this may be due to the very simple measurement error model employed for this variable. In addition, the information on liabilities is generally easier to recall and less sensitive than the information on the assets. This usually results in general in a lower measurement error.

Table 9. Number of households (millions) holding a financial instrument using different estimators

Instrument	$\hat{\mathbf{t}}_{\tilde{y},\,SHIW}$	$\hat{\mathbf{t}}_{\tilde{y},\,NR}$	$\hat{\mathbf{t}}_{\hat{y},\,ME}$	$\hat{\mathbf{t}}_{\hat{y},\,NRME}$	$\frac{\hat{\mathbf{t}}_{\hat{y},\,NRME}}{\hat{\mathbf{t}}_{\tilde{y},\,SHIW}}$
Bonds	2.605	2.877	10.710	10.709	4.11
Risky assets	3.332	3.674	13.184	13.112	3.94
Total financial assets	4.832	5.310	16.021	16.068	3.33
Financial liabilities	5.646	5.881	6.518	6.800	1.20

Table 10. Total amount held (billions of euros) using different estimators, estimate from NFAs, and ratio between survey estimates and NFAs (percentages).

Instrument	$\hat{\mathbf{t}}_{\tilde{y},\, SHIW}$	*% of NFAs*	$\hat{\mathbf{t}}_{\tilde{y},\, NR}$	*% of NFAs*	$\hat{\mathbf{t}}_{\hat{y},\, ME}$	*% of NFAs*	$\hat{\mathbf{t}}_{\hat{y},\, NRME}$	*% of NFAs*	NFAs*
Bonds	102.3	13.5	127.3	16.8	486.2	64.1	501.8	66.1	758.8
Risky assets	126.8	11.5	155.2	14.1	994.3	90.3	1,005.2	91.3	1,100.6
Total financial assets	229.1	12.3	282.5	15.2	1,480.5	79.6	1,507.0	81.0	1,859.4
Financial liabilities	257.6	41.2	298.7	47.8	304.8	48.8	350.8	56.2	624.8

* The figures exclude: cash, bank and postal deposits, and technical insurances.

In order to give a sense of the magnitude of the impact of nonresponse and measurement error, it is useful to compare it with the magnitude of sampling errors: the relative standard error for the total financial assets and for financial liabilities is about 5 and 6 percent, respectively. These figures are negligible compared to the ones shown in the aforementioned tables. The main implication is that surveys on households' wealth require data producers to pay more attention to nonsampling errors rather than to the sampling error alone. Our results are likely driven by the specific features of the SHIW. Yet, when it comes to households' wealth, sampling errors are likely to play a negligible role compared to nonsampling errors. Data producers should therefore allocate their limited financial resources accordingly.

Moreover, in the case of the SHIW, the bias due to measurement error outweighs by far the bias due to unit nonresponse. This result is in part due to how the survey on nonrespondents is designed. The response rate for this survey is very low and there is likely a severe issue of self-selection underneath the sampling of nonrespondents. Therefore, the response model for non-panel households seems to be more a model for the propensity to response over that of providing at least a short interview, rather than over a refusal. The survey on nonrespondents seems to fail to provide information on the real refusing households. For this reason the adjustment for unit nonresponse should be considered as a lower bound.

Finally, measurement error has been dealt with by imputation using model estimates coming from an external validation sample. It would certainly be of interest to investigate the properties of an estimator based on a subsample of households on which an accurate measurement of the variables of interest can be taken. More study is required to determine the sub-sampling design and dimension. Finally, like with imputation for item nonresponse, a fully weighting or an imputation approach can be used to determine the final estimates. Both approaches have pros and cons. The former requires the dissemination of a different set of weights for each variable of interest for which accurate measurements are taken. The latter, on the other hand, allows the computation of a single set of weights, but requires the dissemination of imputed values for units not in the subsample.

Acknowledgement

The work of Ranalli is supported by the Research of National Interest n. 2007RHFBB3PRIN awarded by the Italian government to the University of Perugia. The views expressed in the paper are solely those of the authors and do not necessarily correspond to those of the *Banca d'Italia*.

REFERENCES

Bagozzi, R. (1994). Measuring in market research: basic principal of questionnaire design. *Principles of Marketing Research in: Blackwell Business.*

Biancotti, C., G. D'Alessio, and A. Neri (2008). Measurement error in the Bank of Italy's survey of household income and wealth. *Review of Income and Wealth,* 54-3.

Biemer, P. (2010). Total survey error: design, implementation, and evaluation. *Public Opinion Quarterly* 74, 817-848.

D'Alessio, G. and I. Faiella (2002). Nonresponse behaviour in the Bank of Italy's survey of household income and wealth. *Temi di Discussione del Servizio Studi*, Banca d'Italia 462.

Durrant, G. B. and C. Skinner (2006). Using missing data methods to correct for measurement error in a distribution function. *Survey Methodology,* 32-1, 25-36.

Ekholm, A. and S. Laaksonen (1991). Weighting via response modeling in the Finnish Household Budget Survey. *Journal of Official Statistics* 7, 325-337.

European Central Bank (2009). Survey data on household finance and consumption. ECB, *Occasional paper series* (100).

Folsom, R. E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. *In ASA Proceedings of the Social Statistics Section*, pp. 197-202.

Fuller, W. A., M. M. Loughin, and H. D. Baker (1994). Regression weighting in the presence of nonresponse with application to the 1987-1988 Nationwide Food Consumption Survey. *Survey Methodology* 20, 75-85.

Kim, J. K. and J. J. Kim (2007). Nonresponse weighting adjustment using estimated response probability. *The Canadian Journal of Statistics* 35 (4), 501-514.

Kott, P. S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology* 32 (2), 133-142.

Laaksonen S. and R. Chambers (2006) Survey estimation under informative nonresponse with follow-up, *Journal of Official Statistics* 22, 81-95

Lessler, J. T. and W. D. Kalsbeek (1992). *Nonsampling Error in Surveys*. John Wiley & Sons.

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

Moore, J., L. Stinson, and E. Welniak (2000). Income measurement error in surveys: A review. *Journal of Official Statistics* 16, 331-361.

Ruppert, D., M. P.Wand, and R. Carroll (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge, New York.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. Springer, Berlin, New York.

Wu, C. and Y. Luan (2003). Optimal calibration estimators under two-phase sampling. *Journal of Official Statistics* 19, 119-131.