THE SPANISH SURVEY OF HOUSEHOLD FINANCES (EFF): DESCRIPTION AND METHODS OF THE 2002 WAVE

2004

Olympia Bover

Documentos Ocasionales

N.º 0409

BANCO DE **ESPAÑA**

THE SPANISH SURVEY OF HOUSEHOLD FINANCES (EFF): DESCRIPTION AND METHODS OF THE 2002 WAVE

THE SPANISH SURVEY OF HOUSEHOLD FINANCES (EFF): DESCRIPTION AND METHODS OF THE 2002 WAVE*

Olympia Bover

DG ECONOMICS, RESEARCH, AND STATISTICS, BANCO DE ESPAÑA

(*) The EFF would not have been possible without the willingness to contribute to the project of the Spanish Statistics Office and the National Tax Office. On the technical aspects the work of some of their people should be mentioned, in particular Carmen Anido, Carlos de la Fuente, Juan Luis Pérez, Juana Porras, Jorge Saralegui, and Carmen Ureña. The example of the US SCF and the continuous willingness to help of Arthur Kennickell has also been crucial. We also benefited from the generous account by Giovanni d'Alessio of the Bank of Italy experience. In a prominent position, Cristina Barceló and Pilar Velilla, as part of the EFF team, dedicated much time and effort to the project. We are also very grateful to Manuel Arellano for his continuous support and advice all along the project. The field work company rose to the challenge thanks to the determination of Chelo Perera, Amelia Rochela, Lourdes Vázquez, José Ignacio Wert, and all the interviewers. Last but not least, the cooperation of the responding households is very gratefully acknowledged.

Documentos Ocasionales. N.º 0409
2004

The Occasional Paper Series seeks to disseminate work conducted at the Banco de España, in the performance of its functions, that may be of general interest.

ISSN: 1696-2222 (print)
ISSN: 1696-2230 (on line)
Depósito legal: M. 51919-2004
Imprenta del Banco de España

Abstract

This document describes the making of the first wave of the Spanish Survey of Household Finances. This survey is the only statistical source in Spain by which it is possible to link incomes, assets, debts and consumption at the household level. First we present the sample design, explaining the way oversampling of the rich was performed. Strata were constructed on the basis of the wealth tax, through a blind system of collaboration between the National Statistics Office and the Tax Office. This system, while preserving stringent tax confidentiality requirements, was able to maintain a unique sampling population frame. Second we report survey non response rates and the weight adjustments performed. Finally, we explain the multiple imputation method used to handle item non-response.

Keywords: wealth survey, oversampling of the rich, imputation methods
JEL classification: C81, D31

1 Introduction

Knowledge of the financial behaviour and situation of households is important for economic analysis and the conduct of economic and regulatory policy, and is becoming increasingly relevant given the growing complexity and variety of these decisions. In Spain there is information on the total amounts of household financial wealth and debts (but not on their real assets) through the aggregate Financial Accounts ('Cuentas Financieras'). However, this information is too aggregate to allow the study of important aspects of household finances.

Indeed, aggregate levels are not enough to assess the situation of different types of families and we need to know the distribution of real and financial assets of households, their debts, and their relationship with other variables. For example, it would be very relevant to know how the recent upsurge in household debt is distributed among different types of families. Does the apparent aggregate comfortable situation hide significant fragility for certain households, or are investments and debts held by families able to absorb adverse fluctuations in the markets? Are the households that contracted debt in the past few years the same ones that acquired assets? To mention other examples, until now it was not possible to know which types of families have pension plans, or how young families finance the purchase of their house. For many other relevant questions information at the household level is needed as well.

As a consequence, the Banco de España decided in 2001 to start a survey on household finances ['Encuesta Financiera de las Familias' (EFF)], as it is carried out in other countries. In particular, the Banca d'Italia Survey ['Survey on Household Income and Wealth' (SHIW)] and the US Board of Governors Survey of Consumer Finances (SCF) provided helpful examples.

One distinctive characteristic of the EFF, following the example of the SCF, is that there is oversampling of wealthy households. The distribution of wealth is heavily skewed and moreover some types of assets are held only by a small fraction of the population. Therefore it was judged important to have a sample that would be not only representative of the population but also of aggregate wealth and that would also facilitate the study of financial behaviour at the top of the wealth distribution. This oversampling was achieved thanks to the collaboration of the Tax Office and the Statistics Office.

The EFF is the only statistical source in Spain that allows the linking of incomes, assets, debts, and consumption at the household level. Often macro comparisons are made from ratios of aggregate debt payments (or total debt) and aggregate disposable income (or assets). The EFF will enable the calculation of genuine measures of aggregate debt ratios as means or medians of the individual household ratios. In general a ratio of aggregate magnitudes is different from an aggregate of ratios, but this difference may be specially relevant in the case of asymmetric and long tailed distributions, as it is the case of the magnitudes of interest here.

The purpose of this document is to describe the main features of the making of the EFF. Section 2 describes the questionnaire. In Section 3 the sample design is discussed. Section 4 presents the various steps involved in the fieldwork and provides an analysis of unit non-response. Section 5 describes how sample weights are adjusted to take into account non-response, and briefly discusses other correction strategies. Finally, Section 6 addresses item non-response and presents the imputation methods used.

2 Questionnaire

Contents

The questionnaire is divided into nine main sections which reflect the aims of the EFF. These are as follows:

1. Demographics[1]
2. Real assets and their associated debts
3. Other debts
4. Financial assets
5. Pension plans and insurances
6. Labour market situation and labour income (for all household members)
7. Non-labour income in previous calendar year (2001)
8. Means of payments
9. Consumption and savings

When designing the EFF questionnaire, the example of wealth surveys questionnaires from other countries (in particular, Italy, Netherlands, and the US) were important inputs that were adapted to suit the Spanish situation. For example, debt information is one of the main motivations for the EFF and therefore, along the lines of the SCF, more attention on debt is devoted as compared to the Italian SHIW. On the other hand, more in line with SHIW, there is a less detailed asset categorization in the EFF as compared to the SCF. Moreover, in Spain (like in Italy but unlike the US), account should be taken of extended families living together. This increases the number of questions on labour status and income as well as the complexity of household level questions like the total amount held by the household in some type of bank accounts.

Finally, in contrast to the SCF, we decided to add non-durable and food expenditure questions at the end of the questionnaire given the interest of the relationship between consumption, income and various types of wealth [see Browning et al. (2002), on recommended formulations for these questions and some assessment of their performance].

One paramount consideration all along the making of the questionnaire was to try and keep the length of the interview to an hour on average. This is significantly shorter than in the SCF but is widely regarded as the maximum bearable length in Europe. In fact, after a pilot survey was run showing an average length of 1h 25mn per interview, some questions were dropped. In Table 1 we report figures about the number of questions faced by households and provide some comparisons with the SCF. The median number of questions posed in the SCF is over twice the number in the EFF while in the case of the number of questions posed in euros (i.e. monetary amounts) the SCF median is 1.6 the EFF median.

1. The demographic questions were worded in a way to be comparable with similar questions in other household surveys carried out by the National Statistics Office.

Other points

It was decided to have ‘Computer Assisted Personal Interviews’ (CAPI). Aside from avoiding the punching-in data stage, the use of the CAPI questionnaire helped interviewers follow the correct routes in what is a complex questionnaire and allowed the programming of some basic checks to detect, in the presence of the respondent, potential inconsistencies or errors in their answers. Furthermore, conversions from pesetas to euros (and vice versa) were easily incorporated. These conversions were very much needed to facilitate and check the answers since the field work started when only few months had elapsed since the introduction of the euro.

The field work spanned over seven months, from October 2002 to May 2003. Few questions were added for interviews conducted from January 2003 onwards, to control for potential increases in labour income linked purely to inflation.

3 Designing the sample

Following the example of the US Board Survey of Consumer Finances (SCF) we felt it was important to oversample wealthy households. Indeed, there are many assets held by only a small fraction of the population and therefore a standard random sample would contain too few or no observations at all for many of the relevant analysis one would like to perform with wealth microdata. Thanks to the collaboration of the Statistics Office and the Tax Office it was possible to obtain a sample with a significant oversampling of the high wealth households.

Basis for the oversampling of the wealthy

In Spain there is a wealth tax ('Impuesto sobre el Patrimonio') and it is on the individual wealth tax files information that the EFF oversampling is based. This is in contrast with the SCF where a wealth index is constructed by integrating the information about asset income coming from the individual income tax files since there is no wealth tax in the US. People liable to the wealth tax in Spain were, in 1999 (which was the tax year used in selecting our sample), those with taxable wealth over 104,000 €. In 1999 around 980,000 individuals (corresponding to approximately 700,000 households) filed a wealth tax return.

The choice of defining the wealth strata was based on the SCF intervals and on the households percentile distribution of the wealth tax for Spain. We defined eight strata which were oversampled progressively at higher rates. The intervals are shown in Table 2. Strata 2 and 3 capture slightly less than half of the distribution of taxable wealth. Strata 4, 5 and 6 capture the third quartile except for the last percentile and a half approximately which is represented by the last two strata.

Finally, in Navarre and the Basque Country there was no oversampling of the wealthy because the national Tax Office does not hold the personal tax file information for those regions.

Confidentiality guaranties

The Tax Office is subject to very stringent confidentiality requirements and cannot release, even to the Statistics Office, any personal tax information (not even in the form of intervals). To overcome the problem and enable wealth tax oversampling while preserving confidentiality, the National Tax Office volunteered to actually do the random sample selection herself following the sample design requirements, as instructed by the Bank of Spain and the National Statistics Office.

Thanks to the collaboration of both the Statistics Office and the Tax Office there is a unique population frame for the sampling[2]. The population frame for the sample was the Continuous Municipal Census dated mid-2001, where the units are the households as defined by their address. With this information sent by the Statistics Office to the Tax Office, the Tax Office constructed for each address three variables based on information from both the wealth and the income tax. These data were the starting point for the sampling.

The first variable, the wealth stratum indicator, is based on total declared taxable wealth for the household, which was obtained by adding up the returns of all its members when applicable. The second one, for those filing income tax but not wealth tax, is a variable indicating to which quartile in the national taxable income distribution the household belongs. Finally, information on the per capita income of the household was also added. The income variables were helpful in the selection of sample replacements (as we shall see below), and to ensure that households from all income levels were selected into the sample. The latter was

2. Unlike the SCF where two samples are used (one random and another with oversampling).

obtained by using systematic sampling with random start in a properly ordered data frame. Furthermore, the income quartile indicator was used to correct for non-response in large cities. The tax information available at the time was dated 1999. This entailed some limited mismatch between the two sources.

Sampling

The sampling design was different for the following three cases:

(i) municipalities with more than 100,000 inhabitants. For large towns, the sampling was random within the eight wealth strata.

(ii) municipalities with 100,000 inhabitants or less. For small municipalities, the sampling was a two stage cluster design, with the primary sampling units (PSU or 'secciones censales') being selected first with probability proportional to their population. Further, within PSU the selection of households was different according to the number of wealth tax filers in the PSU.

(iii) Navarre and the Basque Country. Finally, in Navarre and the Basque Country where, as we have mentioned, no oversampling of the wealthy was possible, the sample was selected according to a two stage stratified cluster design with six strata defined according to municipality size.

Replacements

Another relevant aspect of the EFF sample design was the replacement scheme chosen. To try and preserve the oversampling scheme as much as possible, tightly controlled replacements were chosen. An important reason in our case for having controlled replacements was the fact that we do not have any indication of the wealth stratum to which the sample households belong so no 'directed' effort could be applied during the field work were we to discover that the response rate of certain strata was being particularly low.

In particular, up to four replacements were provided for each household originally in the sample that would serve as replacements of that household only. Those replacements were selected to be the two households immediately before and the two immediately after the household in a file ranked by income quartile (for non wealth tax filers), wealth stratum, and per capita household income. Replacements had to belong to the same income quartile (for non wealth tax payers) or the same wealth stratum as the sample household. This was done within municipalities in the case of large cities and within PSU in the case of small ones to keep replacements geographically not too distant from the original sample household. These implied that in some cases less than four replacements were available (and in a few instances, none at all). In the case of Navarre and the Basque country a more standard scheme of a pool of eight replacement households being potential substitutes for eight sample households (within the same PSU) was adopted.

4 The fieldwork

One of the characteristics of wealth and income surveys is high unit non-response due to the nature or the difficulty of the questions asked. In this section we describe the situation for the EFF and in the next one we deal with the corrections adopted to minimize resulting self-selection problems in the sample.

After the fieldwork, the total number of valid completed interviews is 5,143. The number of contacts and attempted contacts underlying this figure is shown in Table 3.

Efforts to reduce non-response

For this first wave we had sent to the households a pack with an introductory letter from the Governor of the Banco de España, another by the field work agency, and a brochure. The letter from the Governor was sent in a separate envelope but inside the one sent by the field work agency to emphasize the point that the Banco ignores the identity and address of the sample households. The objective of this point was to make clear to respondents that their answers to the survey will not be used for fiscal purposes. However, too often those letters were lost in the post or discarded by the household because they looked unimportant. Given these problems, the importance of the letter getting to be read by households may outweight the cost of the marginal suspicion and sending directly an official letter could be more effective. A web page and telephone numbers were also provided to reassure households about the legitimacy of the survey and answer questions they may have. The Banco de España local branches were notified of the ongoing survey and this proved important since many people turned to them for confirmation. Finally, a token gift was offered to participating families.

It was decided to start the field work in early October, after the summer holidays period, because the probability of not finding households at their primary residence is very high from mid-June to mid-September.

Training of the interviewers

There were three one day train-the-trainers sessions in different locations (Madrid, Barcelona, and Bilbao). Those sessions were attended by the interviewers coordinators and some interviewers. Previously a training session with some interviewers' coordinators and interviewers from Madrid had taken place for the pilot survey. During these sessions the questionnaire was analysed by going through hypothetical cases and getting familiar with this particular CAPI application. A representative of the Banco de España participated at these sessions to explain the importance and difficulty of the project and to clarify all matters that were to arise during the explanation of the questionnaire. Later, sessions to train the interviewers were conducted all over the country by the local coordinators. A manual for the interviewers was handed to every interviewer with detailed explanations about the questionnaire, definitions of the concepts involved and examples.

We were aware of the importance of the training of the interviewers in order to minimize non-response. However, training sessions lasting more than one day were unseen in Spain (as well as long interviewers manuals as the one for the EFF) and the budget increase entailed were we to insist on the issue would have been unaffordable. The feeling at the end of those sessions was that more time to train the interviewers would be useful or, alternatively, that more professional training by some expert in surveys of this kind could be very beneficial.

Not possible to establish contact (never at home)

The number of households for which the interviewer was unable to find anybody at home (having confirmed with neighbours etc… that the address corresponds to the household) is very high despite at least five attempted visits. The number of these failed contacts as a proportion of the total number of attempted contacts by wealth strata has some non-random component as we can see in Table 4[3]. Multiple residences was perceived as a potential reason for failing to establish contact with high wealth people during the field work. Overall, the difficulty in getting completed interviews (both because of not being able to contact the household and refusals to participate) convinced the field work agency to increase twice the price per completed interview and to change the incentive scheme of the interviewers by including a fixed per diem stipend in the most difficult areas. Unfortunately this was after some good interviewers dropped from the project.

Refusal

As we can see in Table 4, there is a clear non-random component in cooperation rates [defined as completed/(completed+refused)], decreasing as we move up the wealth strata, ranging from 53.6% to 29.4%. It is clear from this pattern that overall cooperation or response rates are not very informative in case of oversampling since they are dependant on the degree of such oversampling. For some meaningful comparison, we constructed cooperation rates by strata for the 1992 SCF[4]. There cooperation rates for the list sample ranged from 52.6% for stratum 1 to 20.1% for stratum 7.

Supervision and discarded interviews

All the completed interviews were first revised by the field work agency supervisors. Around 67% of the completed interviews were re-contacted (mostly by phone but some personally). There were various reasons for re-contacting: (i) check potential inconsistencies, (ii) confirm all extreme values, and (iii) reduce item non-response.

A program was developed to detect logical inconsistencies between questions. Households sometimes provided a plausible explanation for them. For example in some cases the reason why the reference person in the household appeared as born after their main residence was bought was due to having inherited that residence. However in many cases this was useful to detect errors. This program was not built in the CAPI questionnaire because it would have meant going backwards and forwards in the questionnaire during the interview. In contrast, coding for the detection of basic inconsistencies or errors concerning individual questions was naturally built into the CAPI program.

Aside from the previous reasons, there was also extensive random re-contact to further control the work of the interviewers.

The EFF team at the Banco de España also examined the completed interviews for overall individual coherency. As a result of this process it was decided to discard completed interviews where less than 30% of the questions in euros faced by each individual household were answered and interviews where no income information was provided (neither labour income nor asset income nor assistance income of some kind)[5]. The final number of discarded interviews is shown in Table 3.

3. Figures in Table 4 were provided by the Tax Office.
4. For 1992 sufficiently disaggregated level information on the type of response to attempted interviews was available [in Kennickell and McManus (1993)]. We found cooperation rates a more meaningful and comparable magnitude given the different sample design (in particular, the absence of replacements in the SCF) and the information available than usually available overall response rates.
5. 30% seemed a natural cutting point after having reviewed the informational content of the completed interviews.

Some descriptive analysis of unit non-response

We do not have much information at our disposal about non-participating households. Nevertheless we briefly explore the relevance of this information in explaining non-response and report the results as a descriptive device. We know the province and the size of the municipality for respondents and non-respondents. Furthermore, interviewers were asked to answer four general questions about the household for both the participating households and the ones refusing to do so[6]. These are: (i) type of dwelling, (ii) type of neighbourhood, (iii) type of building, and (iv) social status.

We report in Table 5 a variance decomposition analysis on the basis of a linear probability model including as geographic factors the 48 provinces and size of municipality dummies, and as socio-economic factors the type of dwelling, type of neighbourhood, and social status[7]. A goodness of fit measure for the logit estimation is also provided. As we can see the socio-economic variables reported by the interviewers do not add much to the geographical factors. The interviewers are used to filling these variables for other surveys but they nevertheless lack much explanatory power in this case. Overall the geographical factors and the interviewers information do not seem to explain much of the cooperation decision. Indeed the pseudo-R^2 is 11% and over 30% of the possible pairs in our sample have predicted probabilities that do not rank accordingly to observed responses. We expect they would be even less relevant in an analysis conditional on the sampling frame variables. As we shall see later, corrections for non-response are solely based on the differential non-response by income quartiles, wealth strata and geographical factors performed by the Tax Office when calculating the weights. Further adjustment by regions was considered but not undertaken because of too small cells.

In Table 6 we present logit parameter estimates of the accepted vs. refused decision. For the sake of clarity dummies for the 17 regions instead of the 48 provinces are included. Households in Asturias, Madrid, and Andalusia are the ones with the higher probability of cooperation (59% predicted probability at the mode characteristics for Asturias, 55% for Madrid)[8]. In contrast, households in Cantabria and Galicia are the less likely to cooperate, with predicted probabilities of 23 and 24% respectively. The effect of municipality size is as expected, with people increasingly more likely to refuse as the size of the municipality increases. For example, in the Madrid region, the probability of cooperating increases to 67% for municipalities between ten and fifty thousand inhabitants.

The degree of oversampling in the final sample

Finally, in what follows we give some figures about the degree of oversampling in our final sample. These were kindly provided by the Tax Office due to the confidentiality restrictions. Overall, slightly over 40% of the households that completed the interview correspond to wealth tax filers. This is in line with the SCF where between 35 and 37% of completed interviews come from the list sample. Furthermore, aggregate tax returns information indicates that four per thousand of the population of households hold 40% of total taxable wealth. We would therefore expect to have at most 20 of such households in a 5,000 random sample, an upper bound since it assumes non-differential rate of response. In contrast, our sample contains over 500 of them.

---

6. Note that this information was not recorded for 247 non-participating households that were initially marked as 'other reasons for non-participating' but were coded as refusals after re-coding this 'others' category. For obvious reasons this information is not available either when no contact was established with the household.

7. We also tried adding the type of building variable but it does not make any change to the results and is a variable difficult to interpret.

8. The mode characteristics are: municipality between 100,000 and 500,000 inhabitants, high quality dwelling, average area, medium-medium social status.

5 Correcting for unit non-response and weights

Weights

The sample design weight for each household comes naturally from our unique population frame as the inverse of the probability of being included in the sample. The reference Census population used is dated last quarter of 2002. In a first step these initial weights have been adjusted for non-response within the cells defined by the various sampling frame variables (which are different according to the municipality size). In particular these include: municipality size, wealth strata, income quartiles for non wealth tax filers in large cities, proportion of wealth tax filers in each PSU, size of PSU. Additional adjustment by regions within those cells was not possible because of insufficient sample size. In line with the confidentiality restrictions mentioned before, these first design and non-response weights were calculated by the Tax Office following detailed instructions by the Statistics Office.

Based on the previous weights, an analysis of the estimates from the sample of various population characteristics was undertaken by the Statistics Office. The characteristics considered were: age, education, labour status, sex, and household size. The estimates obtained for education and labour status are satisfactory but the sample was found to be biased towards older individuals and households of smaller size than the population. Therefore the first weights were adjusted (by a linear distance function using the Calmar procedure) to conform to the Census age, household size and sex population structures[9].

In order to improve the calculation of weights, we took into account the age structure by municipality size as provided by the Population Census. We distinguished between large municipalities (greater than 100,000 inhabitants) and the rest. We would expect that the relative response rates between age groups vary by municipality size. This is specially so if we take into account the fact that oversampling by wealth was mainly conducted in large municipalities and that the relative response rates between age groups are likely to differ by wealth.

Further potential corrections for non-response

From the above, the fundamental working assumption is that conditional on the sampling frame variables non-respondents are missing at random [see Rubin (1976) or Little and Rubin (1987)]. In general, the missing at random assumption can be written as $f(Y \mid X) = f(Y \mid X, D=1)$ where Y is an outcome variable of interest, X is the vector of the sampling frame variables and $D=1$ indicates participation into the survey.

Suppose that some additional variables Z are available and that further conditioning provides a more plausible assumption of missing at random $f(Y \mid X, Z) = f(Y \mid X, Z, D=1)$. In that case the weights (evaluated on the basis of X) would have to be multiplied by $Prob(D=1 \mid X, Z) / Prob(D=1 \mid X)$, for example by $\Phi(\alpha + \beta' X + \gamma' Z) \Big/ \frac{1}{N}\sum_{l=1}^{N} \Phi(\alpha + \beta' X + \gamma' Z_l)$, if $Prob(D=1 \mid X, Z)$ is modelled using a probit specification[10].

These corrections for non-response are based on assumptions similar to the ones made in the matching literature where, conditional on X and Z, D is assigned at random. They differ from the Heckman-λ corrections made in the classical selection model. In that

9. Details on the Calmar procedure, developped by the French INSEE, can be found in Sautory (1993). One useful feature of this procedure is that it allows for different levels of adjustment simultaneously, in particular, households and individuals.

10. Monitoring that further weight adjustments do not entail large increases in volatility would be needed.

context D depends on unobservables that are not independent of X [11]. This classical solution is more ambitious since it does not require conditional exogeneity of D. However it is more restrictive because it relies on imposing a specific structural model.

For the EFF no extra valuable information is available with sufficient sample support to explain non-participation. In the SCF an estimated non-response propensity score is available from the rich information on participants and non participants in the list sample. However it is reported to increase the volatility of the weights while not adding much information to adjustments by wealth strata, location, and a measure of financial income [Kennickell (2000)].

11. We have $E(Y \mid X, Z) = X\beta$ where $E(Y \mid X, Z) \neq E(Y \mid X, Z, D = 1)$ and $E(Y \mid X, Z, D = 1) = X\beta + \sigma\lambda(X\delta + Z\mu)$ where $Y = X\beta + u, D = 1$ if $X\delta + Z\mu + v$.

6 Imputation

6.1 Item non-response

Item non-response occurs when a household agrees to participate in the survey but fails to respond to one or more questions. Together with high unit non-response, item non-response is an inherent characteristic of wealth surveys. Moreover they are closely related. Indeed item non-response will partly depend on the stringency of the conditions imposed (in terms of the amount of important questions having to be completed) to declare valid an interview which in turn affects unit non-response rates. This is an issue one has often to address at early stages since it may affect the terms of the contract with the field agency. In particular, there is a trade-off because stringent conditions would give the right incentives to the interviewers but would produce self-selection into the sample in addition to the one created by overall refusals to participate. Moreover, faced with too stringent conditions the interviewers are more likely to cheat or to induce answers from the household.

Answers to the questions on whether the household holds a particular asset are usually readily provided. In contrast, households may have experience more difficulties in providing information about the value of the asset held or about the amount of a particular income source. When designing the questionnaire we considered the possibility of collecting range data in a systematic way by means of successive open range questions in the cases of No Answer/Don't Know to euro questions, as it is done in the SCF or the US Health and Retirement Survey. After some discussions we felt that prompting households with a battery of open range questions might alienate respondents since a priori we suspected that most NA/DK after being confronted with a range card would indeed reflect refusals to provide information. Along these lines, the evidence provided for the SCF about reporting rates by type of information provided (i.e. precise value, card or range) indicates that the information obtained using open ranges is very small [see tables in Kennickell (2000)]. Therefore only range cards were provided to the interviewers to help respondents give information. However into the field work we realized that range cards were almost not used by interviewers on the grounds that respondents were not making use of them. In any case, to avoid this happening because of discouraged interviewers, a way to enforce probing by range cards would be desirable.

In Table 1 some statistics about the number of questions faced and answered per household are provided. More specifically, in Table 7 we present non-response rates to some key questions. For a comparison we look at similar tables for the 1995 and 1998 SCF in Kennickell (2000). The EFF response rates concerning various asset values are pretty close to the ones obtained by the SCF (adding up for the later rates for precise values and range cards). On the other hand response rates for various types of labour related income are substantially higher in comparison for the EFF. In absolute terms rates of response for income from financial assets is low but we do not have comparable information for the SCF.

6.2 Imputation methods and their motivation[12]

Why impute

Given the item non-response rates reported above, working with only the available cases ignoring item non-response would not be sensible. First, this would assume that the complete cases are a random subsample of the original sample. This is most probably not valid (as we have seen for example in the case of unit non-response), and therefore such an analysis could induce severe biases in the results. Second in multivariate analyses, working with only the observations for which all the variables of interest are completed would lead to far too small samples.

• *Imputation for enabling the analysis of the EFF with complete-data methods*. Correct inferences from an incomplete data set can be made using for example model based maximum likelihood methods. However this is not technically available to all potential users of the data. Therefore it is beneficial to provide users of the data with some imputation of (i.e. 'filling in') the missing data, which of course analysts are free to ignore[13].

Imputation is not meant to create artificial information or to give the impression that the data set contains more information than it actually has, but to exploit exhaustively the existing one in a way to enable the various possible analyses of the data using complete data tools.

• *Imputation as a responsibility of the data provider.* Imputation is a resources consuming process which is not at the disposal of most users and is sensibly thought to be the data provider's responsibility [see Rubin (1996)]. An additional reason, very relevant in the case of the EFF, for the Banco de España to provide imputation is that we have access to some information (like some stratifying and location variables) relevant for imputing sensible values which will not be available in the public data file for confidentiality reasons.

Choice of imputation method

Before explaining our choice of imputation method we should say that they all rely on the *missing at random* (MAR) assumption [as defined in Rubin (1976) and in Little and Rubin (1987)]. This requires that the missing values behave like a random sample of all values but within groups defined by observed data. The goodness of this assumption will depend on the availability of observed variables which could plausibly explain missingness and conditional upon which the analysis can be conducted.

One of the central motivations for launching the EFF was to learn about the distribution of the real and financial assets of households, their debts and their relationship with other variables. To preserve the observed distribution of variables and the covariances between them, stochastic imputation methods should be used. Indeed, simple methods like mean imputation (conditional or unconditional) tend to produce peaked distributions of the variables and underestimation of the variances.

A very popular method of stochastic imputation is *hot deck*, with some variations. In general with a hot deck procedure the missing item for a given household would be replaced by the value of the item reported by some similar-in-characteristics household. However in the case of the EFF the number of characteristics/variables upon which one would like to condition before being sensible to assume that the missing information is missing at random is too large to produce reasonably sized cells from which to draw the hot deck imputation. Therefore, most of the EFF imputations, as we will see later, are based on random regression type models.

---

12. The references for this sub-section (except the last part) are Little and Rubin (1987), Rubin (1987), and Schafer (1997).

13. All imputed values are flagged accordingly.

Finally, whatever the stochastic method chosen, one problem with providing one single imputation (i.e. filling in one value for each missing value) is that the imputed values are considered as if they had really been observed. This ignores uncertainty about the imputation under the considered model and additional potential uncertainty when more than one model could be chosen for imputation. Therefore standard errors and other uncertainty measures would be downward biased. Providing multiple imputations (MI), as proposed by Rubin (1987) is hence a desirable goal in the imputation process.

The idea behind multiple imputation is that for each missing value, several imputed values (say *m*) are provided instead of just one[14]. Such a multiply imputed data set gives rise to *m* complete data sets. The way to use the MI data set is (i) first analyze the *m* imputed datasets separately using complete data tools, and (ii) two, combine the results. A concern we initially had is that multiple imputed data could discourage potential users of the data. Indeed, the steps one has to follow to use a MI dataset are in principle simple. Nevertheless, the combination of results might discourage the less technical user for whom in part the imputation process is undertaken, defeating then part of its purpose. However, nowadays, widely used software like Stata can be easily made to perform such tasks [see Carlin et al. (2003)].

Software used for imputation

We have been very fortunate to be allowed to use the programs written at the US Board by Arthur Kennickell [see Kennickell (1991 and 1998)] for the SCF multiple imputation, as well as to benefit from his advice. These programs are specially suited to our case since both surveys share important features that cannot be accommodated in other available packages. In particular, in such complex datasets almost each observation has a different pattern of item missingness. This is more difficult to deal with than monotone patterns where there is a ranking of the variables according to their missingness which applies for all observations. Another important feature available in the SCF imputation programs is the possibility to impose unit specific constraints on the values to be imputed.

The SCF multiple imputation program (Fritz)

In what follows we shall briefly outline the main features of the SCF multiple imputation program[15].

The SCF multiple imputation program (Fritz, Federal Reserve Imputation Technique Zeta) has a sequential and iterative structure. In a given iteration the variables are imputed sequentially and an imputed variable is taken as 'observed' for subsequent imputations in the sequence and in the next iterations (but subject to updating). The order of imputation of the variables in the sequence matters and one consideration is that variables with not many missing cases and relevant as sufficient statistics for predicting other variables are to be imputed first. This iterative and sequential imputation is related to some of the Markov Chain Monte Carlo developments (MCMC), in particular Gibbs sampling.

There are three types of imputations allowed in the Fritz program: continuous, binary, and multinomial. For continuous variables, the imputations are randomizations from regression predictions. An important feature of the program is that it allows to specify all the variables that one would like to use as regressors. However, in the first iteration, instead of using only the coefficients from a regression based only on complete cases, for each variable and observation to be imputed, the program determines the non-missing variables among the full set of regressors and uses the corresponding subset of rows and columns of the

14. Usually *m* between 2 and 10 is sufficient. For example [see Schafer (1997)] with 50% of the information missing, an estimate based on *m*=5 will tend to have a standard error only 1.049 times as large as the estimate with $m=\infty$.
15. More details can be found in Kennickell (1991, 1998, and 2000) and Kennickell and McManus (1994).

covariance matrix necessary for that 'individual' regression. In later iterations the full covariance matrix is available using imputed data from the previous iteration.

For binary variables a variant of the linear probability model is used, to take advantage of the covariance based 'individual' regression feature described. Finally, for multinomial variables a type of hot deck procedure that allows for conditioning on one discrete and one continuous variable at most is used[16].

Some practical issues

• *Evaluation of the imputations.* The set of variables used in the regression should be as general as possible. It should include all possible determinants of the variable to be imputed as well as, in the Fritz system, predictors of these determinants, should they in turn be missing. Furthermore, any variable potentially related to the missingness pattern should be included for the MAR hypothesis to hold. The willingness to have a model as complete as possible has however to be balanced in practice with the limitation imposed by degrees of freedom.

It is not straightforward how to assess the goodness of the imputed values for a given variable. Firstly, no unique goodness of fit statistic of the regression model used in the imputation is available even for the sample of observed values of the variable, given the almost 'individual' regression used for each missing case (as explained above). Moreover, a good within sample fit (i.e. for observed values) need not be necessary nor sufficient for convincing out-of-sample (i.e. non-respondents) imputations.

Secondly, to judge whether imputed variables are 'reasonable', one would like to be able to compare the distribution of the observed values of the variable (Y^O, say) with the distribution of the imputed values (Y^I, say). However, the relevant distributions for a proper comparison should be conditional on the variables (at least on few of them) used in the imputation regression. The problem with such comparisons of conditional distributions is the lack of sufficient observations.

Given these considerations we have implemented nearest neighbours procedures to check for outlying imputations[17]. For each imputed value Y_i^I with associated covariate vector $\underline{X_i}$ we define a set of neighbour observations of respondents (J) as those for which the vector of observed covariates are close to $\underline{X_i}$ and compare Y_i^I with the set of neighbours' Y_J^O. Formally, we define neighbours as those respondents with $\left\| \underline{X_i} - \underline{X_j} \right\| \le \varepsilon$ for a suitably chosen ε [18]. The norm we use is

$$\left\| \underline{X_i} - \underline{X_j} \right\| = \sqrt{(X_{i1} - X_{j1},..., X_{ik} - X_{jk})\Omega^{-1}(X_{i1} - X_{j1},..., X_{ik} - X_{jk})'}$$

where Ω is the covariance matrix of the corresponding covariates. Once a set of neighbours is selected, we check whether Y_i^I lies between the maximum Y_J^O plus one residual standard deviation (and the minimum Y_J^O minus one standard deviation). In case this is not satisfied, the imputed value is reset to the neighbours' maximum (or minimum).

If the number of neighbours found is too small for a robust evaluation the process is repeated including among the set of potential neighbours not only the respondents but also the evaluated imputations[19,20].

---

16. The continuous variable may be defined as the result of the interaction of more than one variable, allowing in this way for some broader conditioning.

17. After the first iteration, to have sensible starting values, and after the last one.

18. We choose a value of ε equal to the square root of the dimension of $\underline{X_i}$ except when the full set of covariates is observed. In such a case we can easily distinguish between discrete and continuous covariates and set ε to $\varepsilon = \sqrt{\text{nº of discrete variables}} + \sqrt{\text{nº of continuous variables}}$.

19. We take the minimum necessary number of neighbours to be 15.

20. Eventually, in the fifth and last neighbours comparison, ε is increased by a tenth of the total number of variables.

• *Convergence of the iterative process.* Another practical concern is how to control for convergence of the imputations after a number of iterations. To that end, in the EFF, for each variable Y for which some values have been imputed we compare the estimations of its median and its interquartile range resulting from successive iterations. In particular, the criterion we use is

$$\sqrt{(Q50_t^Y - Q50_{t-1}^Y, (Q75-Q25)_t^Y - (Q75-Q25)_{t-1}^Y)(Q50_t^Y - Q50_{t-1}^Y, (Q75-Q25)_t^Y - (Q75-Q25)_{t-1}^Y)'}$$

where t is the iteration number, and $Q25, Q50, Q75$ the 25th, 50th, and 75th percentile, respectively. We perform this analysis for the unconditional distribution of each Y and conditioning on some stratifying variables as well.

Since randomization and multiple imputation occur at each iteration, convergence is defined with respect to measures of position and dispersion instead of a point-wise criterion.

REFERENCES

BROWNING, M., T. F. CROSSLEY, and G. WEBER (2002). *Asking Consumption Questions in General Purpose Surveys.*

CARLIN, J. B., N. LI, P. GREENWOOD, and C. COFFEY (2003). 'Tools for analyzing multiple imputed datasets', *The Stata Journal,* 3, pp. 226-244.

KENNICKELL, A. (1991). *Imputation of the 1989 Survey of Consumer Finances: Stochastic Relaxation and Multiple Imputation.*

—(1998). *Multiple Imputation in the Survey of Consumer Finances.*

—(2000). *Wealth Measurement in the Survey of Consumer Finances: Methodology and Directions for Future Research.*

KENNICKELL, A., and D. McMANUS (1993). *Sampling for Household Financial Characteristics Using Frame Information on Past Income.*

—(1994). *Multiple Imputation of the 1983 and 1989 waves of the SCF.*

LITTLE, R. J. A., and D. B. RUBIN (1987). *Statistical Analysis with Missing Data*, Wiley, New York.

RUBIN D. B. (1976). 'Inference and Missing Data', *Biometrika,* 63, pp. 581-592.

— (1987). *Multiple Imputation for Nonresponse in Surveys,* Wiley.

— (1996). 'Multiple Imputation After 18+ Years', *Journal of the American Statistical Association,* 91, pp. 473-489.

SAUTORY, O. (1993). *La macro Calmar. Redressement d'un échantillon par calage sur marges,* Document No. F9310, Institut National de la Statistique et des Études Économiques, Paris.

SCHAFER, J. L. (1997). *Analysis of Incomplete Multivariate Data,* Chapman & Hall, London.

Table 1. Quantity of questions faced and answered per sample household, unweighted

	Average	Median	Standard Deviation	Minimum	Maximum
Nº of questions posed	209	206[1]	56.1	100	462[2]
Nº of € questions posed	23.4	22[3]	9.5	5	73[4]
Nº of questions answered	201.2	198	54.1	91	425
Nº of € questions answered	19.4	18	8.7	3	66
% of questions posed being answered	96.2	96.7	2.4	75.7	100
% of € questions posed being answered	82.7	85.7	14.4	30.8	100

The following are from Kennickell (2000):

1. 434 in the 1998 SCF
2. 784 in the 1998 SCF
3. 35 in the 1998 SCF
4. 107 in the 1998 SCF

Table 2. Wealth strata definition

Stratum 1	do not file wealth tax
Stratum 2	less than 20 million pesetas[1] (120,200 €)
Stratum 3	20 to 50 millions pesetas (120,200 to 300,500 €)
Stratum 4	50 to 100 millions pesetas (300,500 to 601,000 €)
Stratum 5	100 to 200 millions pesetas (601,000 to 1,202,200 €)
Stratum 6	200 to 500 millions pesetas (1,202,200 to 3,005,500 €)
Stratum 7	500 to 2,000 millions pesetas (3,005,500 to 12,022,000 €)
Stratum 8	over 2,000 millions pesetas (12,022,000 €)

1. The definition of the intervals are in pesetas because that was the currency of the 1999 tax records, previous to the introduction of the euro.

Table 3. Number of attempted contacts, by type of response

Completed	5,143
Refused	5,722
Never at home	6,670
Out of scope (wrong address, not a housing unit, empty dwelling, deceased, others out of scope)	1,797
Discarded after supervision	569
Total	19,901

Table 4. Some measures of non-participation, by wealth strata

	Never at home[1]	Cooperation rate[2]
Total	33.5%	47.3%
Stratum 1	31%	53.6%
Stratum 2	38.9%	45.3%
Stratum 3	32.9%	44.7%
Stratum 4	35.5%	46.5%
Stratum 5	37%	38.5%
Stratum 6	38%	36.1%
Stratum 7	40.1%	37.8%
Stratum 8	39.8%	29.4%
Navarre and Basque Country	26%	46%

1. Defined as (Never at home/Total attempted contacts)
2. Defined as (Completed/Completed+Refused)

Table 5. Variance decomposition analysis and goodness of fit measure

Dependent variable = 1 if completed, 0 if refused Number of observations = 10,586				
	Mean	Geographical factors	Socio-economic factors	Geographical and socio-economic factors
Standard deviation	0.50	0.48	0.49	0.47
R^2		0.09	0.03	0.11
Association of predicted probability and observed responses[1]:				
Concordant		65.3%	49.3%	68.5%
Tied		3.1%	18.7%	0.8%

1. As a measure of the goodness of fit we look at the association of predicted probabilities and observed responses. This measures how many pairs of observations have a concordant response, i.e. how many pairs (with our sample we have almost 28 million pairs) with different observed responses have predicted probabilities that rank accordingly.

Table 6. Logit parameter estimates of the completed vs. refused decision[1]

	Coefficient	Odds Ratio	t-ratio
Type of dwelling			
Luxury	-0.4423742	0.6425092	-2.04
Semi-luxury	-0.8676473	0.4199384	-3.68
High-quality	-1.2784430	0.2784705	-5.33
Low-quality	-1.5481060	0.2126504	-6.09
Poor dwelling	-1.2162740	0.2963321	-3.75
Type of area			
High standing	-0.3941852	0.6742292	-2.00
Above average	0.1242046	1.1322470	0.60
Average	0.5946163	1.8123350	2.78
Below average	0.6797912	1.9734660	2.96
Low	1.0055460	2.7333990	3.66
Very low	1.2245340	3.4025810	2.50
Poor	-0.7514802	0.4716679	-1.76
Social Status			
Medium-high	0.2801777	1.3233650	2.45
Medium-medium	0.1091127	1.1152880	0.83
Medium-low	0.5073223	1.6608380	3.39
Low	1.0015720	2.7225590	4.61
Size of municipality			
2,000<inhab=<10,000	-0.7049064	0.4941548	-5.21
10,000<inhab=<50,000	-0.7412587	0.4765137	-5.63
50,000<inhab=<100,000	-0.9232242	0.3972362	-6.61
100,000<inhab=<500,000	-1.2313290	0.2919044	-9.60
500,000<inhab=<1,000,000	-1.4425380	0.2363273	-9.90
inhab>1,000,000	-1.3295370	0.2645998	-9.44
Region			
Aragon	-0.7990338	0.4497633	-6.54
Asturias	0.1853343	1.2036210	1.33
Balearic Islands	-0.7270669	0.4833245	-4.21
Canary Islands	-0.4181743	0.6582475	-3.53
Cantabria	-1.3713790	0.2537567	-9.46
Castille-La Mancha	-0.2665401	0.7660253	-1.92
Castille-Leon	-0.2544959	0.7753073	-2.36
Catalonia	-0.7376957	0.4782146	-8.84
Valencia	-0.4092502	0.6641480	-4.90
Extremadura	-0.2212221	0.8015386	-1.26
Galicia	-1.3394490	0.2619900	-12.74
Madrid	0.0187715	1.0189490	0.19
Murcia	-0.0833291	0.9200483	-0.53
Navarre	-1.2323730	0.2915997	-5.79
Basque Country	-0.5408224	0.5822692	-4.81
La Rioja	-0.4666641	0.6270907	-2.14
Constant	1.9994390		9.42

1. The omitted categories are: very luxury dwelling, very high standing neighbourhood, high social status, municipalities with 2,000 inhabitants or less, Andalusia.

Table 7. Reporting rates (%) of various items, unweighted sample

	Have item		Value for those having the item		
	Yes	Unknown	Value	DK	NA
Own main residence	84.5	0.0	86.5	13.0	0.5
Amount owed, 1rst loan, main residence	15.0	0.0	88.6	11.2	0.3
Monthly payment, 1rst loan, main residence	15.0	0.0	96.2	3.5	0.1
Rent main residence	9.9	0.0	97.4	1.0	1.6
Other real estate, 1rst property	41.7	0.0	82.0	16.4	1.0
Amount owed, 1rst loan, 1rst other real estate	5.0	0.0	91.1	6.6	0.8
Accounts usable for payments	96.9	1.5	74.3	11.7	14.0
Accounts not usable for payments	20.8	2.2	81.8	6.5	11.8
Listed shares	20.7	0.3	76.6	15.9	7.4
Unlisted shares	6.9	0.2	51.3	34.6	14.2
Mutual funds, 1rst fund	14.7	0.2	76.6	12.8	7.5
Fixed income securities	3.3	0.2	81.4	11.0	7.6
Pension plans, 1rst plan	25.8	0.0	62.3	34.6	3.0
Life insurance (1rst policy) coverage	8.9	0.0	63.9	33.5	2.6
Business market value (reference person)	13.1	0.0	64.3	32.3	3.4
Wage income (reference person, 2001)	36.9	0.0	97.6	1.2	1.3
Self-employment income (ref. person, 2001)	13.4	0.0	89.6	5.2	5.2
Unemployment benefits (ref. person, 2001)	1.5	0.0	94.7	5.3	0.0
Pensions (reference person, 2001)	31.8	0.0	99.2	0.2	0.6
Income from real assets (2001)	11.1	0.1	92.0	3.3	4.7
Income from dividends, coupons, etc (2001)	9.3	0.9	60.7	33.4	5.9
Bank accounts interest income (2001)	65.1	3.6	34.1	60.5	5.4
Food expenditure	100.0	0.0	93.8	5.8	0.4
Non-durable expenditure	100.0	0.0	95.9	3.6	0.5

BANCO DE ESPAÑA PUBLICATIONS

OCCASIONAL PAPERS

BANCO DE ESPAÑA	Unidad de Publicaciones Alcalá, 522; 28027 Madrid Teléfono +34 91 338 6363. Fax +34 91 338 6488 *e-mail:* Publicaciones@bde.es www.bde.es