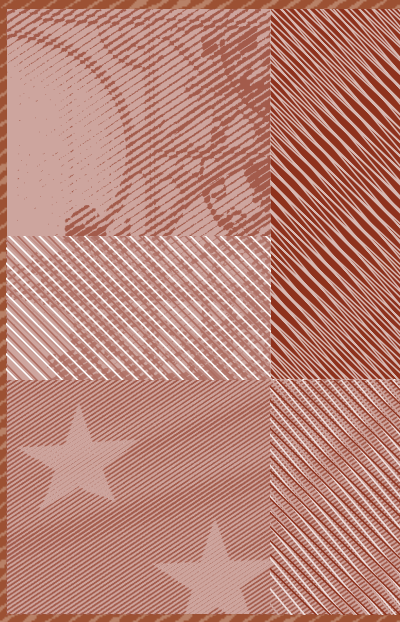
THE SPANISH SURVEY OF HOUSEHOLD FINANCES (EFF): DESCRIPTION AND METHODS OF THE 2005 WAVE

2008

Olympia Bover



Documentos Ocasionales
N.º 0803

BANCO DE ESPAÑA
Eurosistema

THE SPANISH SURVEY OF HOUSEHOLD FINANCES (EFF): DESCRIPTION AND METHODS OF THE 2005 WAVE

THE SPANISH SURVEY OF HOUSEHOLD FINANCES (EFF): DESCRIPTION AND METHODS OF THE 2005 WAVE

Olympia Bover

BANCO DE ESPAÑA

Documentos Ocasionales. N.º 0803
2008

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.

The opinions and analyses in the Occasional Paper Series are the responsibility of the authors and, therefore, do not necessarily coincide with those of the Banco de España or the Eurosystem.

The Banco de España disseminates its main reports and most of its publications via the INTERNET at the following website: http://www.bde.es.

ISSN: 1696-2222 (print)
ISSN: 1696-2230 (on line)
Depósito legal: M.23624-2008
Unidad de Publicaciones, Banco de España

Abstract

This paper describes the methods of the second wave of the Spanish Survey of Household Finances (EFF2005), paying special attention to the innovations relative to the first wave. The EFF2005 was designed to give continuity to the information on household finances collected through the EFF2002. The EFF is the only statistical source in Spain that allows the linking of incomes, assets, debts, and consumption at the household level. A desirable characteristic present in both waves is the oversampling of wealthy households. This is achieved on the basis of the wealth tax through a blind system of collaboration between the National Statistics Institute and the Tax Office which preserves stringent tax confidentiality. An additional important characteristic of the EFF is that the second wave has a full panel component. Further, a refreshment sample by wealth stratum has been incorporated to preserve cross-sectional representativity and overall sample size.

Keywords: wealth survey, oversampling of the rich, panel, refreshment sample, imputation.

JEL classification: C81, D31.

1 Introduction

The Banco de España started conducting the Spanish Survey of Household Finances (EFF) in 2002. The EFF2002 was the only statistical source in Spain that allowed the linking of incomes, assets, debts, and consumption at the household level. The availability of the EFF enabled, for example, the calculation of individual debt ratios, which not only provide a picture of their heterogeneity across households, but also show that they can be quite different to aggregate ratios, in view of the skewed distributions of income, wealth and debt. As another example, the availability of the EFF made it possible to show that the household response of consumption to changes in housing wealth depends not only on the level of wealth but also on characteristics like age[1].

The second wave of the Spanish Survey of Household Finances (EFF), relating to end-2005, was conducted to bring up to date the information on household finances first collected through the EFF2002. It contains the same type of information three years later and hence allows the changes that have occurred since then to be assessed. It also provides an updated representative picture of the structure of household assets and debts at the household level[2]. Moreover, since part of the 2005 sample is a panel, the combined 2002 and 2005 samples provide information on the distribution of individual changes between the two periods.

A desirable characteristic present in both waves of the EFF is the oversampling of wealthy households. The distribution of wealth is heavily skewed and moreover some types of assets are held by only a small fraction of the population. Therefore, unless one is prepared to collect very large samples, oversampling is important to achieve not only representativeness of the population but also of aggregate wealth. Furthermore, it is necessary to enable the study of financial behaviour at the top of the wealth distribution.

Oversampling in the EFF is achieved thanks to the collaboration of the Tax Office and the National Statistics Institute on the basis of individual wealth tax records, while preserving stringent tax confidentiality.

An additional important characteristic of the EFF is that the second wave has a full panel component. Having a panel allows the study of transitions and to account for heterogeneity among households. This was judged important both for descriptive and research purposes. On the other hand, a complete fresh cross-section would be better for capturing the structure of the 2005 population. The compromise adopted in the EFF2005 was to try to re-interview all the EFF2002 respondents and, additionally, to incorporate a refreshment sample to preserve cross-sectional representativity and overall sample size.

This paper describes the main features of the methods of the EFF2005 as well as the innovations that the panel component entails in the design of the EFF2005. Section 2 briefly outlines the questionnaire. Section 3 describes the sample design. Section 4 presents the fieldwork and an analysis of unit non-response. Section 5 describes the final sample, in particular the panel component and the amount of oversampling. Section 6 discusses the calculation of cross-sectional and longitudinal weights. Lastly, Section 7 presents an analysis of item non-response and concludes with some brief comments on imputation issues.

---

1. See Bover (2005).

2. The third wave will foreseeably take place at the end of 2008. The possibility of changing the frequency of the survey to two years is envisaged.

2 The questionnaire

2.1 Contents

The questionnaire is fundamentally the same as in the first wave. We list here its main sections for completeness:

1. Demographics[3]
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 the previous calendar year (2004)
8. Means of payment
9. Consumption and savings

This questionnaire was presented as a 'Computer Assisted Personal Interview' (CAPI) in both waves. Compared to paper questionnaires the use of CAPI facilitates the task of the interviewers in what is a complex questionnaire, allows some basic checks for errors at the interview stage, and enables automatic conversion from pesetas to euros and vice versa.

The median time taken to complete the questionnaire was around 55 minutes and 95% of the interviews took less than two hours, with the minimum length of interview being 20 minutes. For 1% of the interviews the duration was above 210 minutes[4].

Table 1 reports some figures concerning the number of questions put to households. The number of questions asked is very much in line with that for the EFF2002, with just a couple of additional questions at the mean and median.

2.2 Changes with respect to the 2002 wave

In the EFF2002 euro questions could only be answered by giving a point value estimate or a Don't Know/No Answer (DK/NA) reply. In contrast, for most of the euro questions asked in the 2005 wave respondents can also provide self-reported ranges when unable or unwilling to provide a point estimate, and respondents who are moreover unable/unwilling to provide a self-reported range can choose a predefined range from a list[5].

For some major types of assets and debts, questions are asked at the level of each individual asset or debt. For example, questions concerning outstanding loans associated with the main residence are formulated for each of these loans. However, if the respondent did not know the number of such loans (or assets) in the EFF2002 no overall information was sought. In the EFF2005, overall questions have been added for households that give a DK/NA reply when asked the number of such assets or debts.

---

3. The demographic questions were worded so as to be comparable with similar questions in other household surveys carried out by the National Statistics Institute, the EU-SILC in particular.

4. For such cases we tried to confirm the length with the interviewer since some originally very lengthy interview times were due to the interviewer closing the interview once at home (or at the hotel).

5. A self-reported range is defined by a lower and/or upper bound provided by the household. The possibility of having successive open range questions was not considered since we felt it could alienate respondents.

Aside from these fallback overall questions, very few new questions have been added. Noteworthy is the addition of three questions aimed at obtaining information on credit cards actually used for credit as opposed to paying the balance in full each month, as has traditionally been the case in Spain.

3 Sample design

A fundamental characteristic of the EFF sample is that there is oversampling of the wealthy. This was judged important when designing the EFF since typically a small fraction of the population holds a large share of household wealth and, furthermore, many financial assets are held by only a small fraction of the population. Therefore, a standard random sample would not contain enough observations for many key analyses of wealth microdata. Thanks to the collaboration of the National Statistics Institute and the Tax Office, we were able to achieve a significant oversampling of the wealthy in the EFF.

In this second wave of the EFF we had two objectives. Firstly, we aimed to achieve a sample with the same overall characteristics as in the EFF2002, namely a sample representative of the current population with oversampling of wealthy households following the same criteria as in the first wave. Secondly, we wanted a part of this sample to be a panel by re-interviewing the 5,143 households that took part in the first wave. The panel component provides statistical information on transitions between states and individual changes in magnitudes. Moreover, it facilitates the study of causal effects.

To achieve this goal a refreshment sample by wealth stratum was designed to supplement the panel component up to a total sample size of 7,000 households and to ensure that, when used jointly with the panel, the overall sample would fulfil representativity and oversampling requirements. As a preliminary step for the design of the refreshment component, the wealth (and income) tax information of the panel sample was updated.

3.1 Basis for oversampling of the wealthy

In Spain there is a wealth tax ('Impuesto sobre el Patrimonio'). The EFF oversampling is based on individual wealth tax file information. The people liable to wealth tax in Spain in 2003 (which was the tax year used in selecting our sample) were those with taxable wealth over €108,000. In 2003 around 896,000 individuals (corresponding to approximately 655,000 households) filed a wealth tax return.

The choice of wealth strata was based on the percentile distribution of households filing a wealth tax return. For the EFF2002 we defined eight strata which were oversampled progressively at higher rates. The intervals used for the EFF2005 are the same and are shown in Table 2. In the EFF2005 strata 2 and 3 capture slightly less than half of the distribution of taxable wealth. Strata 4, 5 and 6 capture the third and fourth quartiles except for the last three percentiles, which are represented by the last two strata.

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 these regions.

3.2 Sampling design

The population frame for the sample was the Municipal Register as at April 2005, in which the units are households as defined by their address. With this information sent by the National Statistics Institute to the Tax Office, the Tax Office constructed for each address three variables based on information drawn from both wealth and income tax returns. These data were the starting point for the sampling.

The first variable, the wealth stratum indicator, is based on the total declared taxable wealth of the household, which was obtained by adding up the returns of all its members when applicable. The second, for those filing income tax but not wealth tax returns, is a variable indicating the quartile in the national taxable income distribution to which 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 in the sample. The latter was 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 related to 2003. This entailed some limited mismatch between the two sources.

The sampling design differed depending on municipality size. For all provincial capitals (there are 52 of them) and municipalities over 100,000 inhabitants, fresh oversampling was designed to supplement the panel sample by wealth stratum taking into account the updated wealth strata of panel households. Within each of the eight wealth strata the new sampling was random, closely following the sampling procedure used in the first wave for municipalities of that size.

For municipalities with 100,000 or fewer inhabitants there was no fresh oversampling. The sampling was a two stage cluster design in which the primary sampling units (PSUs or 'secciones censales') were the same as those used in the first wave[6]. Within each PSU, households were randomly selected to supplement the panel households belonging to it, up to an overall number of seven households per PSU. In the first wave oversampling in these type of municipalities was achieved only for PSUs with ten or more wealth tax filers. For these PSUs four wealth tax filers and four non-wealth tax filers had been drawn.

Sampling for Navarre and the Basque Country was similar to that for the group of smaller municipalities but with a finer stratification by municipality size for small municipalities. Specifically, the panel sample was supplemented up to a total of seven households within each of the PSUs used in the EFF2002.

3.3 Confidentiality guarantees

The Tax Office is subject to very stringent confidentiality requirements and cannot release any personal tax information (not even in the form of ranges). To overcome the problem and enable wealth tax oversampling while preserving confidentiality, the National Tax Office volunteered to actually do the random sample selection itself as instructed by the Banco de España and the National Statistics Institute.

3.4 Replacements

To try and preserve as much as possible the oversampling scheme devised for large municipalities and all provincial capitals, tightly controlled replacements were chosen[7]. The use of controlled replacements is similar to post-stratification and weight adjustments within cells when data collection is completed. 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, thus ruling out the possibility of a 'directed' effort

6. In the first wave the PSUs were selected with a probability proportional to their population.

7. In the first wave controlled replacements were also chosen in small municipalities in the case of PSUs with 10 or more wealth tax filers.

during the field work should it be found that the response rate of certain strata was particularly low.

In particular, in large cities and provincial capitals up to four replacements were provided for each original household in the sample that would serve as replacements of that household only. These replacements were the two households immediately before and the two immediately after the household in a list ranked by income quartile (for non-filers of wealth tax), wealth stratum, and per capita household income. Replacements had to belong to the same income quartile (for non-filers of wealth tax returns) or the same wealth stratum as the sample household. This was done within municipalities to keep replacements geographically not too distant from the original sample household. In the case of smaller municipalities, Navarre, and the Basque country, a more standard scheme of a pool of eight replacement households as potential substitutes for eight sample households (within the same PSU) was adopted.

4 Fieldwork

The fieldwork lasted six months, from approximately November 2005 to May 2006[8,9]. During that time 15,662 households were contacted (see Table 3 for more details). The fieldwork for wealth and income surveys is particularly demanding because of high unit non-response due to the nature and difficulty of the questions asked.

4.1 Efforts to reduce non-response

A pack with introductory letters from the Governor of the Banco de España and from the fieldwork agency, and a brochure was sent by the fieldwork company to the sample households. A website and telephone numbers were also provided for households to confirm the legitimacy of the survey and answer questions they might have. The Banco de España local branches were notified of the survey since people often turned to them for confirmation.

When visiting households, interviewers took with them some additional documentation to illustrate the way the data collected were used. In particular they would provide the household with a copy of the Banco de España Economic Bulletin article describing the results of the EFF2002 as well as a selection of articles that appeared in a variety of newspapers following the publication of the EFF2002 results. Finally, a token gift was offered to participating families. Furthermore, panel households received a Banco de España book on Spanish banknotes even if they did not agree to collaborate in the second wave.

4.2 Training the interviewers

To minimise non-response and ensure good quality data, the proper training of interviewers is of paramount importance. Experience with wealth surveys in Spain is limited to the EFF and therefore the Banco de España arranged for the NORC at the University of Chicago, which has conducted the last six waves of the U.S. Survey of Consumer Finances (SCF), to come and explain how the training of SCF interviewers is actually conducted. Four NORC representatives who participated in the SCF met in Madrid for five days with representatives of the Banco de España EFF team and of the EFF fieldwork company to conduct full training of interviewers as done for the SCF. The array of techniques and resources was enlightening and clearly showed the desirability of such training for the EFF.

The actual training of the interviewers for the EFF2005 was two days long and conducted at seven locations in Spain (Barcelona, Bilbao, Gijón, Madrid, Seville, Valencia, and Zaragoza). Previously, a similar training session attended by all regional fieldwork co-ordinators had taken place. During these sessions the questionnaire was analysed in detail by going through hypothetical cases and getting familiar with this particular CAPI application. A representative of the Banco de España participated in these sessions to explain the importance and difficulty of the project and to clarify any matters arising during the explanation of the questionnaire. Arguments to reduce non-cooperation were also discussed as well as appropriate ways of approaching households. A manual for the interviewers was handed to every interviewer with detailed explanations about the questionnaire, definitions of the concepts involved, examples, and reasons for participation in the survey.

---

8. The fieldwork and the programming of the CAPI questionnaire were contracted out to Ipsos.
9. Some delay with the programming of the CAPI questionnaire prevented a preferable early October start.

4.3 Tracing panel households

All addresses of households that participated in the EFF2002 were visited. A household was considered a panel household if at least one of its 2005 members was a member of a participating 2002 household. Sometimes the panel status of the people currently living at the panel addresses was not straightforward from current household members' recollection because members involved in answering in the 2002 wave had left or died. During the visit, and in order to establish the panel nature of current members (and match them to the 2002 members), some automatic comparisons of demographic information about household members between the two waves was performed. However, no use of full names was possible since these are not asked to avoid suspicions given the nature of the data. A more thorough inspection of the panel status of households, their members, and their matching between waves was later carried out by the Banco de España, as reported below. Nevertheless, for households at panel addresses who refused to participate this detailed inspection could obviously not be done and we had to rely on the checks performed by the interviewers only.

A fraction of households interviewed for the EFF2002 were not found at their 2002 address[10]. Some of them had moved but for others there seemed to be a mismatch between addresses in the two waves[11]. Efforts were made to trace and re-interview households that participated in the EFF2002 but had moved since. Overall, 45 of them were interviewed at a new address.

4.4 Never at home and Refusals

As seen in Table 4, aggregate co-operation rates [defined as completed/(completed+refused)] for the whole sample mask significant differences between the panel and the non-panel components[12]. Overall, the co-operation rate of the panel component is 67.1% compared to 38.6% for non-panel. These differences are large in all strata[13]. The smallest differences occur for those not filing wealth tax returns (stratum 1) and for Navarre and the Basque Country (approximately 22 percentage points difference) and the differences are larger for higher strata (37, 34, and 45 percentage points difference for strata 6, 7, and 8, respectively). Furthermore, while there is a clear non-random component in these rates for the non-panel sample (they decrease as we move up the wealth strata from 47.2% to 21.8%), this is not the case for the panel sample.

There is an important difference in the co-operation rates of the non-panel component (38.6% overall) and those obtained in the EFF2002 (47.3%). Moreover, this is true for all strata. These lower co-operation rates in 2005 may be due to insufficient monitoring of the interviewers (in turn probably related to the lack of a satisfactory sample management programme) preventing feedback and adjustment of interviewer practices.

The number of households for which the interviewer was unable to find anybody at home after five attempted visits (having confirmed that the address corresponds to the household) is not large (10.2% of total attempted contacts) and, overall, similar for the panel

---

10. This accounts for the difference of 434 in the total number of panel attempted contacts (4,709) and the EFF2002 sample size (5,143). As noted above, this was more clearly established when the non-panel household living at the panel address agreed to participate as a replacement (356) than when it refused.

11. The interviews that could be held with new households presently living at a panel address showed that in approximately 25% of cases the 2002 household had moved out and that 75% of cases were address mismatches (as calculated using the year in which the current household bought the property, when available -233 cases).

12. The figures in Table 4 were provided by the Tax Office due to confidentiality restrictions.

13. The 90.9% co-operation rate for stratum 2 was calculated from only a small number of panel households contacted in that stratum.

and non-panel parts of the sample. By strata, never at home rates are significantly lower for the top three strata in the panel part as compared to non-panel. Finally, comparison of the EFF2005 and EFF2002 shows that many more addresses were visited without finding anybody at their main residence in 2002 (6,670 of a total number of attempted contacts of 19,901).

As a descriptive device, Table 5 presents logit parameter estimates of the accepted vs. refused decision to participate in the EFF2005, along with some information at our disposal about non-participating households. We separate the panel and the non-panel samples given the very large differences in unconditional co-operation rates just described above. The most noteworthy feature that emerges is that while for non-panel households the probability of co-operating diminishes with municipality size, this probability does not vary significantly with municipality size in the case of panel households. As for other variables, the building condition and type of area variables recorded by the interviewer do not provide very telling results. Regarding differences across regions, households in the Basque Country, Catalonia, and Castille-Leon are significantly less inclined to co-operate as compared to the rest, both in panel and in non-panel households.

4.5 Control and validation

The data from the completed interviews were revised in detail by the team at the Banco de España to uncover potential inconsistencies and implausible values. This process was delayed since the fieldwork company was not able to provide timely data from all the completed interviews while the fieldwork was underway. During the process of revising the data, the EFF team looked at almost 5,000 completed questionnaires. When additional information or clarification was considered important, the fieldwork company recontacted the household. The trade-off between getting additional information and bothering households was taken into account by the EFF team for each individual case. Additional information was obtained for about 1,000 households. The most common errors found in the recorded answers were: (i) monthly vs. annual quantities, (ii) euro vs. pesetas, and (iii) incorrect interpretation of particular questions by some interviewers.

The EFF team at the Banco de España also examined the completed interviews for overall individual consistency. As a result of this process it was decided to discard: (i) completed interviews where no income information was provided (neither labour income nor asset income nor assistance income of any kind), except in the case of panel households with a high percentage of answered euro questions, and (ii) interviews where less than 30% ofthe questions in euro were answered, unless that percentage increased substantially when answers provided in range form were taken into account. These conditions emerged as natural cut-off points after having reviewed the informational content of the completed interviews and are in line with those adopted for the EFF2002. The final number of discarded interviews is shown in Table 3.

As explained above, the actual panel status of households was not always clear cut during the interview and therefore the Banco de España conducted *a posteriori* an additional panel identification process. A procedure was developed to determine the clearly panel (and non-panel) cases, around 560 households were individually revised and their information was compared to the 2002 information to determine whether the household was really panel (or non-panel). Once this was established, further analysis to link specific members of the household between waves was needed and approximately 640 questionnaires were reviewed and compared to the corresponding 2002 information. As a result of the

process it emerged that the panel nature of households was not correctly established during the interview for 239 final sample households.

5 The final sample

5.1 Panel and refreshment in the final sample

The total number of valid completed interviews is 5,962. There are 2,580 households in the EFF2005 sample (43.3%) that had also participated in the EFF2002, i.e. 50% of the EFF2002 sample. Table 6 shows the changes in composition of the panel households between the two waves. In particular, 69.3% of them (i.e. 1,787) have neither gained nor lost members, 7.3% (188) have one additional member, and 17.5% (451) have lost one member. The number of individual household members interviewed in the two waves is 6,744.

There are two different components in the non-panel part of the sample, namely 1,211 households (20.3% of the sample) that replace non-cooperating EFF2002 households and 2,171 (36.4% of the sample) that are refreshment households.

5.2 Degree of oversampling in the final sample

The number of wealth tax filers in the final sample is very similar for the two EFF waves. In the 2005 wave they represent around 35% of the sample[14] while in the population the proportion of household that filed a wealth tax return is around 4.6%.

Regarding actual net worth in the EFF data, Table 7 presents oversampling rates in various parts of the distribution for the two waves[15]. The oversampling rate is defined as the ratio of the number of observations actually in the sample for a specific percentile range of the distribution to the number of observations one would expect if the sample was randomly drawn from the population. As can be seen, a progressive oversampling of the wealthy is achieved. In particular, in the EFF2005, for the wealthier 1% the number of observations is around nine times what would be expected with random sampling.

14. These figures were kindly provided by the Tax Office due to confidentiality restrictions.
15. EFF2005 net worth data correspond to the preliminary imputations dated autumn 2007.

6 Correcting for unit non-response and weights

In the EFF2005 both cross-sectional and longitudinal weights are provided. In line with the confidentiality restrictions mentioned above, design and non-response weights were calculated by the Tax Office following detailed instructions from the National Statistics Institute. In this section we describe the construction of the weights. For details on further potential corrections for non-response and the relationship with econometric selectivity corrections, see Bover (2004).

6.1 Longitudinal weights

The initial weights for the panel households were their 2002 design weights corrected for 2002 non-response. These were further corrected for the non-response in 2005 of the 2002 sample, using as reference the 2002 population. Non-response corrections in both EFF waves are made in the cells defined by the various sampling frame variables. In particular these include municipality size, wealth stratum, and income quartile for non-filers of wealth tax returns.

In a second step, the aforementioned weights were adjusted to conform to the 2005 population, by wealth stratum and income quartile. Finally, these were further adjusted (by a linear distance function using the Calmar procedure) to conform to the 2005 Census structure of the population according to gender, age by municipality size, and household size by municipality size[16] [17].

6.2 Cross-sectional weights

To obtain cross-sectional weights, the panel and non-panel components of the sample are considered as two independent samples.

The basic weights for non-panel households are the inverse of the probability of being included in the sample (as given by the sampling design), subsequently adjusted for non-response within the cells defined by the various sampling frame variables. For panel households, the basic weights are the longitudinal weights prior to their Calmar adjustment, as described earlier.

Finally, the two sample components are combined and their weights corrected according to the relative size of the sub-samples, this being the minimum variance estimator for two independent samples representing the same population. The resulting weights were adjusted using the Calmar procedure to conform to the 2005 Census structure of the population according to gender, age by municipality size, and household size by municipality size.

---

16. Details of the Calmar procedure, developed 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.

17. Another set of longitudinal weights that are adjusted to conform to the 2002 population will also be provided.

7 Item non-response and imputation

7.1 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 (in terms of the number of key questions that have to be completed) that have to be met for an interview to be declared valid, which in turn affects unit non-response rates. This is an issue that often arises in the 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 incentive to interviewers but would produce self-selection into the sample in addition to that created by overall refusals to participate. Moreover, interviewers faced with overly stringent conditions are more likely to cheat or to induce answers from the household. The fieldwork contract conditions in the EFF2005 were the same as in the EFF2002.

The number of questions answered (reported in Table 1) increases somewhat as compared to 2002. In particular, the percentage of euro questions answered (excluding ranges) increases from 85.7% to 91.7% at the median, but the dispersion is substantial. The figures in Table 1 are similar for the panel and non-panel components of the sample.

Answers to the questions on whether the household holds a particular asset are usually readily provided. In contrast, households may have more difficulty providing information about the value of the asset held or about the amount of a particular income source. In the EFF2005 we introduced the possibility that for questions in euro the household could give answers in the form of a range when not able or not willing to provide point values. Namely, when the household answered DK (don't know) to the point value question, he/she was prompted to provide an answer as a self-reported range (as defined by an upper and a lower bound) or, if failing to do so, to chose from a set of predefined ranges.

In Table 1 we document the number of questions answered by the household, distinguishing for the euro questions between answers in point values, self-reported ranges, and predefined ranges from a list. Almost 30% of the sample (1,756 households) gave at least one of their euro answers choosing a predefined range from the list but only 5.6% (333 households) provided self-reported ranges. In any case, range answers was not used extensively, as we can see from the statistics provided. For example, the number of questions answered by a single household in the form of a predefined range was 2 at the median, 2.4 at the mean, and 16 at the maximum. As a percentage of the euro answers provided by a household, these figures would be 8.3%, 11.3% and 66.7%, respectively[18].

At the same, time however, information provided in the form of ranges (and more particularly as predefined ranges) appears to have reduced significantly the proportion of DK/NA answers, mainly the DK ones, without reducing the number of point value responses. This can be seen by comparing the non-response rates to some key questions in Table 8 with a similar table for the EFF2002 [see Bover (2004)]. For example, the percentage of DK answers to the question on the value of the first real estate property (other than the main

18. Percentages not shown in the table.

residence) was 16.4% in the EFF2002, rather similar to the EFF2005 sum of DK (5.0%), predefined range (9.1%), and own range (1.1%). The proportion of NA remains almost unchanged in this case (1.0 in EFF2002 vs. 0.8 in the EFF2005). In contrast, if we look at the response rates of a euro question for which the range option was not introduced in 2005, for example income from dividends, coupons, etc. in the year prior to the interview, it is striking how in that case the proportion of DK is almost the same in both waves (33.4% in 2002 and 34.3% in 2005).

7.2 Imputation methods

In the EFF2005 imputation of DK/NA answers was performed using the same methods as in the EFF2002 (for a general rationale and description see Bover (2004), for a detailed explanation of the procedures and the models involved see Barceló (2006), and for a comparison of the performance of different imputation methods see Barceló (2008)[19]).

However, although the same framework and methods were used, the models for all the variables were revised and often modified as a result of the new data. Moreover, given the newly introduced possibility of range answers, imputation was performed subject to the imputed values belonging to the range provided by the household, when applicable.

The new panel aspect in this second wave of the EFF would in principle allow a new imputation of the 2002 EFF data using the information obtained in 2005, and vice versa. This has not yet been done and the imputations provided so far are static ones. However, forward and backward imputation is an avenue we plan to explore. To get an idea of the amount of information that could be gained from a dynamic imputation in Table 9 we calculate, for some key questions, the conditional probabilities of not giving a point value answer to a euro question in the EFF2005 having provided one in the EFF2002 (and vice versa). These indicate that in general more information might be gained from backward imputation than from forward imputation.

19. In both waves nearest neighbours procedures described in Bover (2004) were implemented only for the first iteration of the imputation process. When preparing the final EFF2002 data this was judged superior to using them in the final imputation as well.

REFERENCES

BARCELÓ, C. (2006). *Imputation of the 2002 wave of the Spanish survey of household finances,* Occasional Paper No. 0603, Banco de España.

— (2008). *The impact of alternative imputation methods on the measurement of income and wealth: evidence from the Spanish survey of household finances,* forthcoming.

BOVER, O. (2004). *The Spanish survey of household finances (EFF): description and methods of the 2002 wave,* Occasional Paper No. 0409, Banco de España.

— (2005). *Wealth effects on consumption: microeconometric estimates from the Spanish survey of household finances*, Working Paper no. 0522, Banco de España, more recent version at: http://www.olympiabover.name/Housing%20wealth_June07.pdf.

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.

Table 1. Number of questions asked and answered per sample household, unweighted

	Average	Median	Standard deviation	Minimum	Maximum
No. of questions asked[1]	211	208	56.4	98	448
No. of € questions asked					
- excl. ranges	26	25	11.1	5	77
- incl. ranges	29	26	13.1	5	94
No. of questions answered[1]	205	202	54.7	90	436
No. of € questions answered					
- point value	23	21	10.5	3	77
- self-reported range[2]	1.5	1	1.0	1	8
- predefined range[3]	2.4	2	2.1	1	16
% of questions answered[1]	97.3	98.1	2.7	77.6	100
% of € questions answered					
- excl. ranges	86.5	91.7	15.1	17.6	100
- incl. ranges	89.5	93.3	12.3	29.4	100

1. Excluding ranges.
2. For those 333 households who provide some answers in self-reported range format.
3. For those 1,756 households who provide some answers choosing a range from the list provided.

Table 2. Definition of wealth strata

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

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

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

	TOTAL	PANEL	NON-PANEL
Completed	5,962	2,580	3,382
Refused	6,634	1,263	5,371
Never at home	1,602	479	1,123
Out of scope (wrong address, not a housing unit, empty dwelling, deceased, others out of scope)	1,275	363	912
Discarded after supervision	189	24	165
Total	15,662	4,709	10,953

Table 4. Some measures of non-participation (%), by wealth stratum

	TOTAL		PANEL		NON-PANEL	
	Never at home[1]	Co-operation rate[2]	Never at home	Co-operation rate	Never at home	Co-operation rate
Total	10.2	47.3	10.2	67.1	10.2	38.6
Stratum 1	8.4	54.9	9.3	68.8	7.9	47.2
Stratum 2	14.3	47.3	12.5	90.9	14.5	42.4
Stratum 3	12.5	46.1	12.6	69.4	12.5	35.3
Stratum 4	11.6	46.9	12.1	67.7	11.3	36.1
Stratum 5	12.5	36.3	13.0	62.1	12.4	29.5
Stratum 6	14.1	35.3	11.2	63.5	14.9	26.5
Stratum 7	14.3	31.4	10.4	58.8	15.2	25.0
Stratum 8	13.0	31.1	10.0	66.7	13.7	21.8
Navarre and Basque Country	5.1	44.3	7.4	61.0	4.3	38.5

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

Table 5. Logit parameter estimates of the completed vs. refused decision[1, 2, 3]: Panel vs. non-panel sample

	Odds ratio	t-ratio	Odds ratio	t-ratio
	Non-panel sample		Panel sample	
Building condition				
Good	0.774	2.81	0.941	0.40
In need of some maintenance	1.054	0.48	0.896	0.60
Very poor	1.769	2.28	1.174	0.47
Type of area				
High standing	0.746	2.39	0.509	2.52
Medium	0.975	0.19	0.518	2.38
Medium-low	1.075	0.51	0.499	2.45
Low	1.134	0.74	0.501	2.19
Size of municipality				
2,000<inhab=<10,000	1.032	0.21	0.747	1.44
10,000<inhab=<50,000	0.946	0.39	0.655	2.18
50,000<inhab=<100,000	0.704	2.36	0.762	1.30
100,000<inhab=<500,000	0.587	3.86	0.568	2.95
500,000<inhab=<1,000,000	0.526	4.03	0.525	2.85
inhab>1,000,000	0.458	5.14	0.663	1.91
Region				
Aragon	0.470	5.10	1.331	1.23
Asturias	0.707	2.35	0.826	0.93
Balearic Islands	1.228	1.17	1.032	0.10
Canary Islands	1.093	0.65	0.997	0.01
Cantabria	0.435	4.06	1.515	1.39
Castille-La Mancha	0.867	0.93	1.021	0.10
Castille-Leon	0.556	5.04	0.525	3.81
Catalonia	0.515	7.19	0.605	3.72
Valencia	1.039	0.39	0.905	0.72
Extremadura	1.260	0.99	1.476	1.38
Galicia	0.676	3.17	0.849	0.85
Madrid	0.835	1.73	0.741	1.96
Murcia	1.491	2.24	1.137	0.51
Navarre	0.504	3.24	1.161	0.35
Basque Country	0.576	4.79	0.565	3.24
La Rioja	1.012	0.05	1.346	0.71
Number of observations	8,916 of which 3,546 yes (39.8%)		3,865 of which 2,604 yes (67.4%)	
Pseudo-R^2	0.04		0.02	

1. The omitted categories are: luxury building, very high standing neighbourhood, high social status, municipalities with 2,000 inhabitants or less, Andalusia.
2. Includes discarded after supervision in the accepted category.
3. Four observations were not included because no information was recorded on the building condition and type of area.

Table 6. Change in the composition of panel households

	No. of members that dropped out between the 2002 and the 2005 wave				Total
	0	1	2	3 or more	
No. of new members in 2005 compared to 2002					
0	1,787	416	97	26	2,326
1	148	24	11	5	188
2	34	8	6	1	49
3 or more	11	3	2	1	17
Total	1,980	451	116	33	2,580

Table 7. Degree of oversampling in the final sample

Net worth decile group	EFF 2002		EFF 2005	
	Number of observations	Oversampling rate[1]	Number of observations	Oversampling rate
Bottom 50%	1,878	0.73	2,234	0.75
50% to 90%	1,944	0.94	2,036	0.85
90% to 95%	429	1.67	481	1.61
95% to 99%	524	2.55	675	2.83
Top 1%	368	7.16	536	8.99

1. The oversampling rate is defined as the ratio of the number of observations actually in the sample for a specific percentile range of the distribution to the number of observations one would expect if the sample was randomly drawn from the population.

Table 8. Reporting rates (%) of various items, unweighted sample

	Have item		Value for those having the item					
	Yes	Un-known	Point value	Own interval	Fixed interval	DK	NA	NP/NF[1]
Own main residence	84.0	0.0	87.3	1.1	7.5	3.5	0.3	0.2
Amount owed, 1st loan, main residence	18.2	0.0	92.4	0.7	4.4	1.9	0.5	0.0
Monthly payment, 1st loan, main residence	18.2	0.0	97.9	0.1	0.8	0.6	0.4	0.0
Rent main residence	9.6	0.0	99.0	n. a.[2]	n. a.	0.5	0.5	0.0
Other real estate, 1st property	48.3	0.0	83.7	1.1	9.1	5.0	0.8	0.0
Amount owed, 1st loan, 1st other real estate	6.1	0.0	91.8	0.0	3.0	3.3	0.5	0.0
Accounts usable for payments	94.4	0.2	79.3	0.1	7.8	3.8	8.8	0.2
Accounts not usable for payments	21.7	0.3	80.6	1.5	6.2	2.9	8.4	0.3
Listed shares	22.5	0.3	79.3	1.4	9.2	6.9	3.0	0.1
Unlisted shares	6.6	0.2	68.9	1.5	9.9	14.5	4.1	1.0
Mutual funds, 1st fund	17.4	0.3	78.1	0.9	6.4	7.7	3.4	0.3
Fixed income securities	2.8	0.4	81.8	1.2	4.7	8.8	3.5	0.0
Pension plans, 1st plan	31.2	0.0	77.4	1.0	8.2	10.0	3.1	0.1
Life insurance (1st policy) coverage	13.5	0.0	73.6	n. a.	n. a.	24.2	1.6	0.6
Business market value (reference person)	12.9	0.0	71.2	1.0	8.5	11.9	7.2	0.3
Wage income (reference person, 2004)	37.9	0.0	91.3	0.2	4.6	0.8	1.2	0.2[3]
Self-employment income (ref. person, 2004)	13.1	0.0	82.0	0.9	9.1	3.6	3.6	0.9
Unemployment benefits (ref. person, 2004)	1.9	0.0	92.0	0.0	7.1	0.0	0.9	0.0
Pensions (reference person, 2004)	30.3	0.0	93.1	0.5	2.2	2.4	1.3	0.4
Income from real assets (2004)	12.9	0.1	91.7	n. a.	n. a.	5.7	1.8	0.8
Income from dividends, coupons, etc (2004)	10.1	1.4	63.2	n. a.	n. a.	34.3	2.3	0.2
Bank accounts interest income (2004)	32.7	3.2	57.4	n. a.	n. a.	40.8	1.8	0.0
Food expenditure	100.0	0.0	96.0	0.2	2.3	1.3	0.2	0.1
Non-durable expenditure	100.0	0.0	96.3	0.3	2.4	0.8	0.1	0.1

1. NP/NF: not plausible/not formulated.
2. n. a.: Answers in ranges were not available for the related question.
3. Additionally 33 cases (1.5%) provided monthly rather than annual amounts and some imputation is needed.

Table 9. Conditional probabilities of not giving a point value answer to a € question in the EFF 2005 having provided one in the EFF 2002 (and vice versa), unweighted panel component of the sample (%)

	Pr (Point value 2005 = 0 | Point value 2002 = 1)	Pr (Point value 2002 = 0 | Point value 2005 = 1)
Own main residence	3.2	11.2
Amount owed, 1st loan, main residence	2.0	8.8
Monthly payment, 1st loan, main residence	0.5	1.9
Rent main residence	0.6	0.6
Other real estate, 1st property	3.4	11.1
Amount owed, 1st loan, 1st other real estate	0.7	0.7
Accounts usable for payments	8.7	21.8
Accounts not usable for payments	3.0	5.7
Listed shares	3.8	12.9
Unlisted shares	6.5	19.5
Mutual funds, 1st fund	4.9	10.0
Fixed income securities	3.7	1.6
Pension plans, 1st plan	5.9	21.9
Life insurance (1st policy) coverage	8.1	7.3
Business market value (reference person)	14.2	25.0
Wage income (reference person, 2004)	0.7	1.1
Self-employment income (ref. person, 2004)	4.8	6.3
Unemployment benefits (ref. person, 2004)	0.0	2.9
Pensions (reference person, 2004)	3.2	0.3
Income from real assets (2004)	6.0	2.1
Income from dividends, coupons, etc (2004)	10.3	10.7
Bank accounts interest income (2004)	17.3	39.8
Food expenditure	1.2	5.8
Non-durable expenditure	0.9	4.0

BANCO DE ESPAÑA PUBLICATIONS

OCCASIONAL PAPERS

0501 JOSÉ RAMÓN MARTÍNEZ-RESANO: Size and heterogeneity matter. A microstructure-based analysis of regulation of secondary markets for government bonds.
0502 ALICIA GARCÍA-HERRERO, SERGIO GAVILÁ AND DANIEL SANTABÁRBARA: China's banking reform: an assessment of its evolution and possible impact.
0503 ANA BUISÁN, DAVID LEARMONTH AND MARÍA SEBASTIÁ BARRIEL: An industry approach to understanding export performance: stylised facts and empirical estimation.
0504 ANA BUISÁN AND FERNANDO RESTOY: Cross-country macroeconometric heterogeneity in EMU.
0505 JOSÉ LUIS MALO DE MOLINA: Una larga fase de expansión de la economía española.
0506 VÍCTOR GARCÍA-VAQUERO AND JORGE MARTÍNEZ: Fiscalidad de la vivienda en España.
0507 JAIME CARUANA: Monetary policy, financial stability and asset prices.
0601 JUAN F. JIMENO, JUAN A. ROJAS AND SERGIO PUENTE: Modelling the impact of aging on Social Security expenditures.
0602 PABLO MARTÍN-ACEÑA: La Banque de France, la BRI et la création du Service des Études de la Banque d'Espagne au début des années 1930. (There is a Spanish version of this edition with the same number.)
0603 CRISTINA BARCELÓ: Imputation of the 2002 wave of the Spanish Survey of Household Finances (EFF).
0604 RAFAEL GÓMEZ AND PABLO HERNÁNDEZ DE COS: The importance of being mature: The effect of demographic maturation on global per-capita income.
0605 JUAN RUIZ AND JOSEP VILARRUBIA: International recycling of petrodolars. (The original Spanish version has the same number.)
0606 ALICIA GARCÍA-HERRERO AND SERGIO GAVILÁ: Posible impacto de Basilea II en los países emergentes.
0607 ESTHER GORDO, JAVIER JAREÑO AND ALBERTO URTASUN: Radiografía del sector de servicios en España.
0608 JUAN AYUSO, ROBERTO BLANCO AND FERNANDO RESTOY: House prices and real interest rates in Spain.
0701 JOSÉ LUIS MALO DE MOLINA: Los principales rasgos y experiencias de la integración de la economía española en la UEM.
0702 ISABEL ARGIMÓN, FRANCISCO DE CASTRO AND ÁNGEL LUIS GÓMEZ: Una simulación de los efectos de la reforma del IRPF sobre la carga impositiva.
0703 YENER ALTUNBAŞ, ALPER KARA AND ADRIAN VAN RIXTEL: Corporate governance and corporate ownership: The investment behaviour of Japanese institutional investors.
0704 ARTURO MACÍAS AND ÁLVARO NASH: Efectos de valoración en la posición de inversión internacional de España.
0705 JUAN ÁNGEL GARCÍA AND ADRIAN VAN RIXTEL: Inflation-linked bonds from a central bank perspective.
0706 JAVIER JAREÑO: Las encuestas de opinión en el análisis coyuntural de la economía española.
0801 MARÍA J. NIETO AND GARRY J. SCHINASI: EU framework for safeguarding financial stability: Towards an analytical benchmark for assessing its effectiveness.
0802 SILVIA IRANZO: Delving into country risk. (There is a Spanish version of this edition with the same number.)
0803 OLYMPIA BOVER: The Spanish survey of household finances (EFF): Description and methods of the 2005 wave.

BANCO DE ESPAÑA
Eurosistema

Unidad de Publicaciones
Alcalá, 522; 28027 Madrid
Telephone +34 91 338 6363. Fax +34 91 338 6488
e-mail: publicaciones@bde.es
www.bde.es