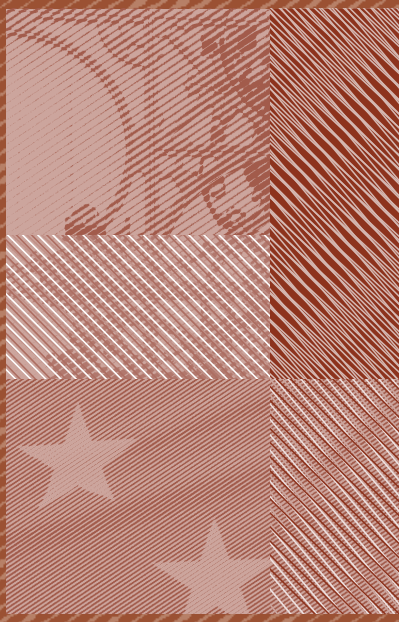
THE SURVEY OF FINANCIAL COMPETENCES (ECF): DESCRIPTION AND METHODS OF THE 2016 WAVE

2019

Olympia Bover, Laura Hospido
and Ernesto Villanueva



Documentos Ocasionales
N.º 1909

BANCO DE **ESPAÑA**
Eurosistema

THE SURVEY OF FINANCIAL COMPETENCES (ECF): DESCRIPTION AND METHODS OF THE 2016 WAVE

THE SURVEY OF FINANCIAL COMPETENCES (ECF): DESCRIPTION AND METHODS OF THE 2016 WAVE(*)

Olympia Bover, Laura Hospido and Ernesto Villanueva

BANCO DE ESPAÑA

(*) We thank Josep Mestre and María Torrado for their commitment and invaluable help during the preparation of the ECF, and Sofía García-Barbadillo, Carlos Sunyer and Carla Veiga for their assistance in different phases of the project. For their help with the implementation of the survey, we thank Inés Andújar, Julio Gil and Fernando Tejada. Finally, for their comments and help with the questionnaire, we thank Adele Atkinson, Cristina Barceló, Vicente Castellanos, Irma Clots, Laura Crespo, Vicente García Rubert, John Gathergood, Pablo Hernández de Cos, María Luengo-Prado, Annamaria Lusardi, Covadonga Sánchez, Pilar Velilla as well as the Financial Education team at the Comisión Nacional del Mercado de Valores (CNMV).

Documentos Ocasionales. N.º 1909

2019

ISSN: 1696-2230 (on-line edition)

Abstract

The Survey of Financial Competences (ECF) is a joint initiative of the Banco de España and the Comisión Nacional del Mercado de Valores (CNMV) aimed at measuring the financial competences of the adult population in Spain. Between 2016 and 2017, information has been collected for a large sample provided by the National Statistical Institute of randomly selected individuals, and representative of the whole Spanish population between 18 and 79 years of age and of each of its regions. This paper provides a detailed description of the most relevant methodological aspects in the design and implementation of the survey: the sample design, the questionnaire, the data collection process, the validation of the data, the computation of weights, and the imputation procedure.

Keywords: financial competences, personal finance, household finance, data collection, imputation, weights.

JEL classification: C81, D14.

Resumen

La Encuesta de Competencias Financieras (ECF) es una iniciativa conjunta del Banco de España y la Comisión Nacional del Mercado de Valores (CNMV) cuyo objetivo es medir las competencias financieras de la población adulta en España. Entre 2016 y 2017 se recogió información para una muestra de gran tamaño, proporcionada por el Instituto Nacional de Estadística, de individuos seleccionados al azar y representativa de toda la población española de entre 18 y 79 años, y de cada una de sus comunidades autónomas. Este artículo proporciona una descripción detallada de los aspectos metodológicos más relevantes del diseño y la implementación de esta encuesta: el diseño muestral, el cuestionario, el proceso de recogida de los datos, la validación de estos, el cálculo de los pesos y la imputación.

Palabras clave: competencias financieras, finanzas personales, finanzas del hogar, recogida de datos, imputación, pesos.

Códigos JEL: C81, D14.

INDEX

1 Introduction

Evidence from a variety of countries suggests that a sizable fraction of the population has deficient knowledge about basic financial notions on inflation, interest rates, or risk diversification. Furthermore, even when one compares individuals with similar degree of education or income, those with lower financial knowledge hold lower wealth levels, higher debt and a higher degree of financial stress.[1] As financial products have become increasingly complex over time, it is important to identify the characteristics and choices of groups of the population with insufficient financial knowledge.

In 2016, the *Banco de España (BdE)* and the Comisión Nacional del Mercado de Valores (CNMV) launched the Survey of Financial Competences (ECF, by its Spanish acronym), an initiative aimed at measuring the financial competences of the population aged 18-79 in Spain. Apart from being representative of the Spanish population as a whole, one important feature of the ECF is the collection of representative information of financial knowledge and practices of the population of each of the 17 Autonomous Communities in Spain (except the Autonomous Cities of Ceuta and Melilla). Representativeness at the regional level is important because the competences about education are transferred to regional authorities. In addition, the socio-economic composition of the population varies widely across Autonomous Communities.

The ECF follows a blueprint questionnaire elaborated by the International Network of Financial Education (INFE), an assessment of the financial knowledge, attitudes and behaviour of representative samples of the adult population. The assessment was recently implemented in a large number of countries under the coordination of the Organization of Economic Cooperation and Development (OECD). The results of the ECF are then comparable to those in other countries.[2] However, the ECF contains a number of extensions and adaptations to the Spanish case that we describe in Section 2.

This article is organised as follows. The second section briefly outlines the questionnaire and describes the adaptations of the blueprint questionnaire to the Spanish case. The third section describes the sample design in detail. The fourth section provides relevant information on some aspects of the data collection process such as the monitoring of fieldwork, the editing of the data and the response rates achieved. The fifth section discusses the weighting. Finally, the sixth section presents an analysis of unit non-response and provides some remarks on the imputation procedures used in selected variables of the survey in cases of item non-response.

1 See, for example, Lusardi and Mitchell (2011), Alessie, Lusardi and Van Rooij (2011) or Stango and Zinman (2009).

2 For a detailed description of the main results of the ECF, see: https://www.bde.es/bde/en/areas/estadis/Otras_estadistic/encuesta-de-comp/ECF_2016.html

2 The questionnaire and the CAPI interview

2.1 Contents

The questionnaire of the ECF follows a blueprint by the International Network on Financial Education (INFE), implemented in a large number of countries in 2015-2016. That exercise measures what financial products individuals know, hold or have recently acquired. If respondents have recently acquired one product, the blueprint also asks which type of information was most influential when choosing that product. Another main block of the INFE exercise is a series of questions measuring financial knowledge. Those questions measure if the respondent is familiar with the notion of interest rate and if they can apply it in simple computations. Further questions measure if respondents understand basic financial concepts such as interest rate compounding, risk diversification and inflation.

The ECF questionnaire applies the INFE measurement exercise. However, it also includes some adaptations to the Spanish context. Firstly, the ECF asks about the respondents' labour market status, individual expectations about employment loss (if employed) or employment finding (if unemployed), as well as measures of impatience. That information is aimed at characterizing precisely the expectations and the attitudes of the individuals interviewed.

A second feature of the ECF is that it is not only a survey on individuals, but also aims at collecting information about the household's level of financial competence. Namely, the survey asks each individual (sample person) if he or she is familiarized with household's asset and debt holdings, as well as about the reasons that determined some financial choices. If the individual is not familiar with the financial matters, an informed person provides complementary information about the household portfolio and his or her own financial competences. Measuring household's competences is especially important in countries where large households with different generations are common.

Thirdly, with the aim of disentangling financial from other generic competences, the ECF also includes items measuring basic reading comprehension (extracted from the 1994 International Adult Literacy Survey) as well as graph comprehension (extracted from the 2012 Program for International Adult Competences, PIAAC).

Finally, the ECF includes a module on housing, given its prominent weight in consumer's portfolios. Following surveys like the National Financial Capability Study in the United States, respondents are asked if they own or rent their main residence, the reasons for their choice and their expectations about the evolution of the price of the dwelling. If respondents financed the purchase of their house of residence with a mortgage still outstanding, the survey collects some information about basic loan characteristics.

In particular, the ECF questionnaire is divided into the following ten main sections:

A Demographics and labour market status
B Portfolio of sample person and means of obtaining information
C Sources of income: inactivity and old age
D Attitudes to saving
E Financial literacy
F Economic literacy
G Household portfolio and means of obtaining information
H Financial literacy
I Main residence
J Expenditure and financial fragility

The first six sections are always asked to the randomly selected person. Sections G and H are asked to an informed household member only when the randomly chosen individual is not familiar with the finances of the household. The information in the last two sections is obtained at the household level, either by the sample person, if he or she is familiar with the finances of the household, or by an informed household member (otherwise).

2.2 Other characteristics

The information was collected by means of personal interviews with the individuals, conducted between end-September 2016 and end-May 2017. The interviews were conducted by interviewers with specific training and were computer-assisted (CAPI). As discussed below, computer-assisted interviews allowed implementing basic consistency checks. In addition, computer-assisted interviews facilitate data handling and allow interviewers to navigate throughout the questionnaire.

The median time taken to complete the ECF questionnaire was around 36 minutes and 90% of the interviews took less than 52 minutes. Only for 4.5% of the interviews was the duration above 1 hour. In cases were two people were asked (6% of the cases), the median duration of the part answered by the sample person was 27 minutes whereas the median duration for the informed person was 25 minutes. Table 1 reports some descriptive figures concerning the number of questions individuals were asked.

3 Sample design

The sample was drawn by the National Statistical Institute from the most recent Census updated with the Municipality Population Registry to be representative of the population of individuals aged 18-79, living in private households in Spain. It is also meant to be representative of each of the 17 regions as, to a large extent the educational system is articulated at that level. The ECF is a probability sample, obtained through a two-stage sampling design, stratified in the first stage units by region and municipality size. The first stage units are the census sections in the country and the second stage are the people between 18 and 79, who have their main residence in the sections selected for the sample.

The reference period was mostly the moment of the interview, but there were exceptions. Some questions referred to the last 12 months (whether household expenses exceeded income, or if the household had fallen into arrears). There were also retrospective questions about the last 2 years (if the individual had acquired a financial product during that period) or the last 5 years (if the individual had had some disagreement with a financial institution). Finally, some questions are about the interviewee's expectations on the price of the main residence during the 12 months following the survey, or expectations about the probability of losing his/her job if working or of finding a new job if unemployed.

4 Fieldwork

The fieldwork was conducted from end-September 2016 to end-May 2017.[3] During this period, 8,952 individuals completed an interview, although after the validation and editing process, 398 interviews were discarded for various reasons (see the last heading of this Section for more details). Table 2 contains the distribution of interviews by month over the fieldwork period, which shows that by the end of December around 78% of the total number of valid interviews were already completed.

Special efforts were devoted to specific strategies designed to minimise non-response and measurement errors such as training of the interviewers, gaining cooperation protocols, and analysis and validation of the data.

4.1 Efforts to mitigate non-response

Each sample person received an envelope containing an introductory letter signed by the Governor of the *Banco de España* and the President of the CNMV. The envelope also contained a letter by the fieldwork agency and a brochure describing the main aims of the study. The introductory letter from *Banco de España* and the CNMV was anonymous and included on a separate envelope within the one sent by the fieldwork company to emphasize the point that neither institution knew the identity and address of the potential interviewee. The letter from both institutions also contained a *Banco de España* webpage and a telephone number to reassure on legitimacy of survey and to answer any question. It was written in every co-official language of each region (Catalan and Spanish in Catalonia and the Balearic Islands, Basque and Spanish in the Basque Country, Galician and Spanish in Galicia and Valencian and Spanish in Valencia). Finally, local branches were informed about the ongoing survey in case sample members turned to them for confirmation.

Contractually, each interviewer had to make at least five attempts to locate each sample member in person (at least once on weekends). However, the number of attempts may exceed that threshold, especially for sample members who had not been reached by the end of the fieldwork (see Table 3).

4.2 Training of interviewers

In mid-September 2016, immediately prior to the start of the fieldwork period, one hundred and nine interviewers went through a centralized two-day training about the contents of the survey, the protocols, and the use of support material (cardboards, letters from *Banco de España* and CNMV, and a small gift for participants). Interviewers were divided into four groups, each trained by a member of the fieldwork company. A representative of *Banco de España* participated in each group to clarify matters that were to arise during the explanation of the questionnaire. A manual for interviewers was handed out, containing definitions of relevant variables, examples and some classifications (like 4-digit occupations).

3 Kantar TNS was selected by *Banco de España* to be in charge of the programming of the CAPI questionnaire and the data collection process.

Aside from issues general to all surveys, the protocol included training points specific to a Survey of Financial Competences. The first training point was the relevance of reading questions exactly as posed in the questionnaire. For example, and unlike financial surveys, when asking if the interviewee knew some specific financial product, interviewers were instructed not to provide definitions–as the survey aims at measuring knowledge of these products. Interviewers were also instructed to detect basic inconsistences and introduce explanatory comments. Such inconsistencies are especially likely to arise in a sample that is representative of individuals aged 18-79 as some respondents delegated financial matters to other household members. For example, some individuals answered that they did not have a bank account, but subsequently mentioned that they held financial products that usually require such product. Those comments, especially common in interviews to the least financially literate respondents, proved very useful during the edition phase.

A third training point stressed the role of the interviewer in imposing discipline when measuring competences. Interviewers asked other family members to leave the room during the section measuring financial knowledge or, if that was not possible, they emphasized that no one could assist the sample member. On the other hand, interviewers were instructed to record if the interviewee used items like paper and pencil or a personal calculator in financial competence questions.

A final training point referred the coding of occupation of the interviewee. That variable was included with the double purpose of obtaining information about the socio-economic status of respondents, as well as about the tasks they conducted in their main job. In particular, respondents had to provide a verbal description of their current or main occupation, which the interviewer coded and subsequently introduced into an automatic search algorithm that proposed the corresponding 4-digit occupation.[4]

By the end of the two-day training, all interviewers took a test, consisting in conducting a shortened interview where one of the members of the training team acted like the interviewee. Interviewers either performing poorly in that test, or coding answers incorrectly underwent additional training before going into the fieldwork or moved to other studies. Eventually, one hundred interviewers proceeded to fieldwork.

4.3 Fieldwork controls

4.3.1 IMMEDIATE FOLLOW-UP AFTER EACH INTERVIEW

Within 48 hours after each interview had been conducted, the supervisors of the fieldwork company phoned the interviewee to check that the guidelines had been implemented. The phone supervisor asked the interviewee the recalled duration of the interview, whether the interviewer asked the consent to be recorded, used a tablet, showed cards or gave a small gift.

4 The intention was to obtain a classification of occupation up to 2 digits (level at which information is publicly released) as accurate as possible.

In addition, the supervisor confirmed selected responses in the interview using a pre-defined script –for example, household composition and housing tenure were confirmed during phone supervision.

In specific cases when the phone supervision was not successful, the fieldwork company sent another interviewer to verify in person that the interview had taken place. *Banco de España* was immediately told about any incidence.

4.3.2 TIME AND AUDIO RECORDING

The time duration of 10% of the questions was recorded, and neither the interviewer nor the interviewee knew which ones.[5]

In addition, interviewers asked all interviewees their consent to be recorded. The wording of the question clarified that the purpose of the audio was to check whether the interview was following the protocol. Upon acceptance, the tablet recorded 5% of the questions, and neither the interviewer nor the interviewee knew which ones. 71% of interviewees gave their consent, and the fraction increased as fieldwork progressed, from 63% in September 2016 to 80% in May 2017.

Audio recording is useful for reasons beyond fieldwork control. It allows a better understanding on how interviewees receive complex questions. A first example is a question about the main occupation of the respondent. There, respondents described the tasks they did at their jobs and their level of responsibility. The interviewer assigned then a 4-digit ISCO code using an automated search tool provided by National Statistical Institute. Audios were very useful in guaranteeing that interviewers were eliciting the information needed to code occupation correctly.

A second example refers to the measurement of expectations about the evolution of the house of residence over the next year. In that question, respondents allocated 10 points to 5 predefined price changes. Audios were again crucial to monitor if the interviewee understood the question correctly.

Duration and audio recording were important monitoring tools that allowed detecting anomalous interview durations or audio recording rates once we used cases in similar locations as a benchmark. Monitoring interviews during the fieldwork was essential to provide quick feedback when errors were detected.

4.4 Refusals and non-contacted

Banco de España required that all individuals were visited in person by interviewers. As an additional requirement, a minimum of 5 in-person contacts distributed among different times

5 The only exception was a question on reading comprehension, where interviewers measured the time required to read and understand a written text unrelated to financial matters.

and days of the week had to be made for each sample member. *Banco de España* and TNS closely monitored the fieldwork process using the data on contacts entered by the interviewers in their case management application. Interviewers were instructed to register detailed information on all contacts and incidences for each individual. Overall, of the original sample of 21,250 individuals, 16,025 individuals were contacted during the fieldwork period.[6] The final data on contacts showed that the average number of in-person visits was 4.25 (the median was 3) and across individuals the percentage of those visits conducted during weekends was 8.45% on average. By type of response, we obtain that valid completed cases received 3.35 in-person visits on average and that 7.31% of these individuals received at least one in-person visit during weekends. Refusal cases received on average 5.91 in-person visits, whereas 9.48% of them received at least one visit during the weekend; while for other non-response they received on average 2.63 in-person visits, 23.30% of them during the weekend. Finally, those cases that were not finally contacted personally received on average 9.78 in-person visits and 15.64% were visited at least once during the weekend.

Table 4 shows different indicators of the fieldwork result based on the final outcome of the attempts of reaching each sample member.

First, the "response rate" indicates the proportion of achieved valid completed interviews in all the eligible cases. Eligibility is defined as the sum of all those valid completed interviews, plus the refusals, other response type and the non-contacted individuals. It is 52.7% in the ECF.

Second, the "contact rate" measures the proportion of sample members contacted by the interviewer, even though they subsequently refused to answer or they were unable to give any type of information. It is calculated by dividing the sum of all the completed valid interviews, plus the refusals, and other non-response type by the eligible population, and it is 98.8% in the ECF.

Third, the "cooperation rate" indicates the proportion of achieved completed valid interviews among those successfully contacted by an interviewer. Thus, it might be considered as a measure of the success in the implementation of gaining cooperation strategies. It is calculated by dividing the sum of all the complete interviews by the sum of all the completed valid interviews, plus the refusals, and other non-response type, and it is 53.4% in the ECF.

Fourth, the "refusal rate" indicates the proportion of individuals that refused to answer in all the eligible cases. It is calculated by dividing the sum of refusals by the eligible population and it is 41.4% in our sample.

Finally, the table also report the "Non-eligibility rate" that indicates the proportion of individuals over the total of 21,250 cases that are out of the sample because of errors in

6 See Table 3 for more details.

frame, empty houses, inaccessible and unreachable addresses, or individuals that have moved with unknown address. It is 23.7% in our sample. This number is around 10 p.p. higher than conventional rates in households' surveys. One reason for the relatively number of high non-eligible individuals in the ECF is that it can be more difficult to reach a particular person with a given name than having any member of a household living in a given address answering a survey (which is in many instances the case in household surveys). Furthermore, during the fieldwork, the fieldwork company reported an unusually high fraction of individuals no longer residing in the addressed specified in the register – either because they had permanently left the country or other reasons. For that reason, at the end of the field, further efforts were made to try to reverse weak refusals and locate sample members not yet located.

To further explore unit non-response, Table 5 presents estimates from a linear regression model for the individual probability of acceptance to participate in the ECF using the information available for all successfully contacted individuals (with not missing information on any control variable considered). In particular, the list of regressors includes measures of the building condition, type of area, municipality size, and region of residence of the interviewee, as well as the gender and age of the sample individual. In addition, interviewers' characteristics collected during their training and selection such as gender, age, educational attainment, number of years of experience working as an interviewer, type of contract, and grades received during the accreditation test are also held constant. We additionally control for the total number of cases assigned to each interviewer.

Regarding location characteristics, the main findings suggest that, overall, the probability of co-operating is higher the better is the building condition of the residence, while it decreases with the size of the municipality. The probability of cooperation is lower for females, and also decreases with the age of the sample individual. Regarding the interviewers' characteristics, keeping the number of assigned cases constant, having higher experience in the field and holding an upper secondary degree are both positively related to cooperation. Finally, grades at training do not seem to be significantly correlated with sample members' cooperation decisions.

4.5 Interviewer incentives and production

In addition to the training, selection and supervision of interviewers, the reward system for interviewers represents another important aspect that should be considered when trying to improve productivity and data quality. In particular, the optimal strategy would be to design an interviewer pay system not only based on response rates and productivity indicators but also on the quality of the data.

Payment per completed case as opposed to fixed weekly/monthly pay is the system used by most survey agencies in Spain. However, given the complexity of the study, it was deemed important for interviewers to earn some fixed pay, despite the fact that such a scheme requires a closer monitoring of personnel by the survey agency. Additionally, and

in order to reward production, the interviewers earned a bonus per interview completed, which varied according to the number of completed interviews they achieved and depending on whether they interview only the sample member or this individual plus another informed person. Interviewers were also aware that they were closely monitored and their interviews fully reviewed and supervised.

In the ECF, 100 interviewers went into the field and 98 of them completed at least one valid interview. The distribution of completed cases among them was as follows: 6 interviewers completed fewer than 10, 35 completed between 10 and 49, 35 completed between 50 and 149, 20 completed between 150 and 249, and 2 interviewers completed more than 250. The median number of interviews completed per interviewer was 57.5 (the mean was 87.3). Twenty-one interviewers completed approximately 50% of the cases in the final sample. Table 6 summarises the main characteristics of the interviewers who went into the fieldwork and obtain at least one valid interview. Specifically, 75.7% of them were females, 69.4% were aged between 36 and 55, around 72% had been working as interviewers for more than 5 years and 34% of them had a permanent contract.

4.6 Control and validation

The CAPI program was developed to detect basic inconsistencies between questions. As already mentioned, those inconsistencies are likely to arise if some individuals are unsure about their own financial situation. The program allowed re-routing in case errors were detected. For example, some individuals had financed high expenses using a personal loan, despite having reported earlier in the survey that no loan was contracted during the last two years. The program automatically re-asked the questions missing – in this case, the source of information used to acquire the loan.

Aside from those checks, there was an extensive validation of interviews during fieldwork, with the aim of detecting errors and providing feedback to interviewers as soon as possible. To that end, the personnel from *Banco de España* had provided a one-day training course to four editors in the fieldwork company. The validation was done via a Web-based platform that permitted access to each (properly anonymized) interview. The field work company and *Banco de España* could both access each interview, but not at the same time. The platform included fields that allowed three rounds of interactions about each case.

The protocol was the following. Once finished, each interview was reviewed by the fieldwork company, which could recommend through the platform changes in the data, re-contacting the interviewed household or, in extreme cases, completely redoing (parts of) the interview. In all of those cases, the *Banco de España* team reviewed the proposals of the fieldwork company and decided on a case-by-case basis. On the other hand, the *Banco de España* team also reviewed at random interviews marked as correct by the fieldwork company to make sure that all interviewers were receiving feedback about their performance.

In cases where some answers were contradictory with previous ones, and after the approval of *Banco de España*, the fieldwork company re-contacted the interviewee by phone. The most common reasons for re-contacting an individual were doubts about his or her degree of financial inclusion. For example, an individual could report not having a bank account, but receiving public pension income. As most of these payments are made through bank accounts, there was a reasonable doubt about whether or not the individual held the financial product. Another common case for re-contact was that of individuals reporting a collateralized loan but not real estate property whatsoever. Finally, inconsistencies about household composition – because the one reported in the interview differed from the one in the phone supervision – or about the labour market status of the individual always led to a phone re-contact.

Finally, in a handful of cases, the *Banco de España* decided that the interview had to be repeated because of deficiencies in measuring financial competences or strong doubts if it had really taken place.[7]

As a result of this validation process, the field work company provided immediate feedback to each interviewer, either positive or providing suggestions for improvement. That feedback was especially important during the first weeks of fieldwork to avoid the accumulation of errors.

The main validation tools were the duration of questions, geo-location of interviews and, especially, audios. Durations served as a basic tool in detecting if some interviewers shortened or read too quickly important questions. On the other hand, audios turned out to play an important role. Firstly, they helped in detecting deviations from the protocol, like imprecise wording or non-neutral attitude of the interviewer – especially relevant in the Sections of financial competences. Secondly, knowing that the interview was recorded helped to discipline the interviewer.

4.7 The final sample

Once interviews had been screened for quality, criteria were established to discard interviews because of insufficient information or doubts about its implementation. The total number of valid interviews completed was 8,554. In less than a quarter of a percent of cases (0.2%), interviews were discarded either because of a refusal to answer section B on financial products heard of, held or recently acquired. About half a percent cases (0.42%) were deemed invalid because the financial competences were incorrectly measured (for example the wording of the questions did not follow the questionnaire literally), and 2% were excluded because of doubts about the protocols followed during the interview (for example, whether a Tablet had been used, or if some questions were skipped during the interview).

Furthermore, in 516 cases, the individual reported not to be informed about the household financial matters and in practically all cases, two persons were interviewed (the sample person and one informed person).

7 In total, all interviews were reviewed by the fieldwork company and about 40% were reviewed additionally by the *Banco de España* team.

5 Correcting for unit non-response and weights

Cross-sectional weights computed by the National Statistics Institute (INE) are provided as part of the data. In this section, we describe the construction of these weights.

The basic weight for each individual is 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.

The resulting weights were adjusted using the Calmar procedure to conform to the most recent structure of the population according to gender, age and nationality at the regional level, based on the 2011 Census (and the Padrón Continuo, a continuously updated municipal population register).

6 Item non-response and imputation

6.1 Item non-response

Item non-response occurs when an individual agrees to participate in the survey but fails to respond to one or more questions. Together with unit non-response, item non-response is an inherent characteristic of surveys. Moreover, they are closely related. Indeed, item non-response will partly depend on the stringency of the conditions that have to be met for an interview to be declared valid (in terms of the number of valid answers), which in turn affects unit non-response rates. As already mentioned in Section 4, in the case of ECF, the contract conditions with the field agency did not set a fixed rule with a minimum number of valid answers below which when an interview would be considered invalid, as such threshold was established once the revisions had been conducted.

Regarding the main patterns of item non-response in the ECF2016, in many instances "Do not know" is a very informative response for questions on knowledge or holding of financial products. A more appropriate definition of non-response would focus on "No answer" type of responses, whose prevalence is testimonial in questions about having heard of, holding or having recently acquired a particular financial product. Similarly, the vast majority of individuals answered questions about demographics, competences (either financial or general), attitudes or expectations.

In contrast, individuals had more difficulty in answering questions on monetary amounts. In anticipation of such concerns, interviewers were instructed to collect information on respondent-provided ranges when the respondent was not able or not willing to provide point monetary values. Furthermore, as some of those questions were recorded, ranges were also introduced during revision. In this case, Table 7 shows the percentage of "Do not know"/"No answer" in questions involving monetary quantities. About 10% of individuals did not provide information about the yearly household income. Item non response fell to 6% in the case of food expenditure and to about 3% when considering expenses in education. In the Section on housing, the fraction of indebted owners not answering to questions about the mortgage instalments or loan-to-value at purchase is about 6%, a percentage that is halved in the question to renters about the monthly rent paid.

6.2 Imputation methods

Imputation is a procedure to provide users with a complete dataset by generating a distribution of m possible values for each missing answer, where each of the values is drawn from the distribution of non-missing information of similar individuals (plus a perturbation term that reflects the uncertainty in the estimation).

As already mentioned, the ECF is a survey that elicits information of what individuals know about their own financial situation as well as about their financial competences.

Hence, “Do not know” responses do not reflect an incomplete dataset, as they are just as informative as complete information on whether or not the individual holds a particular asset, say. To reiterate, the ECF is not a survey designed to capture the financial situation of individuals in detail, so providing a distribution of possible values for each “Do not know”/”No answer” answer is not strictly needed.

However, many users may be interested in conducting analyses of the distribution of financial competences by household income, or to relate quantitative measures of economic fragility to attitudes towards finance, without having to adjust for the fact that not all respondents reported their income or answered all expenditure questions.[8]

For those reasons, a small set of quantitative questions that the respondent was not able or willing to answer was imputed, together with some unreliable responses. In particular, the variables that were imputed were those shown in Table 7, and they include the monthly rent (for individuals who answered that their household rents the dwelling), the monthly mortgage instalment and the loan-to-value ratio at purchase (for homeowners with an outstanding mortgage loan), as well as household expenses in food and education. Similarly, household income, asked in six brackets, was imputed as well. We note that imputation was done exclusively in cases when it was certain that the individual (who didn´t know, didn’t want to reply or wasn’t asked) should have answered the question. That is, m imputations are made for the rental amount paid by renters, but no attempt was done to impute whether the home ownership of an individual who did not answer whether the dwelling is owned or rented (so those variables are left as “not applicable”, following the convention of the survey).

6.2.1 CHOICE OF IMPUTATION METHOD

The imputation methods used in the ECF 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.

The idea behind multiple imputation is that for each missing value, several imputed values (say m) are provided instead of just one. Such a multiply imputed data set gives rise to m complete data sets. The way to use the multiply imputed data set is to first analyze first the m imputed datasets separately using complete data tools, and to combine the results in a second step.

All the ECF imputations are based on random regression type models, performed using STATA Multiple Imputation programs. In a given iteration the variables are imputed sequentially and an imputed variable is taken as ‘observed’ for subsequent imputations in the sequence

8 The advantage of data producers providing imputed values, rather than leaving the task to users, is that the former have access to restricted-use variables with explanatory power in explaining household income or expenses, such as detailed geographical location.

and in the next iterations (but subject to updating). There are three types of imputations in ECF: continuous (for example, expenditure in food), binary (whether an individual had spent anything on education during the last 12 months) or multinomial (income, provided in six brackets, or the distribution of the loan-to-value, that exhibits bunching at particular regulatory values).

For continuous variables, the imputations are randomizations from regression predictions. As shown in Table 7, respondent-provided brackets were available for all questions and, in some cases, the interviewer or reviewer collected both the single point and an interval as an answer. In those cases, the estimated likelihood function included terms modelling the interval where the bracketed response lied as a function of the covariates included.

An important feature of the STATA MICE program is that it allows specifying all the variables that one would like to use as regressors. To fix ideas, assume income has predictive power in explaining the loan-to-value. However, as shown in Table 7, around 10% of respondents do not provide a bracket for the household income. The first iteration predicting the loan-to-value (say) uses the coefficients from a regression based only on cases where the dependent variable is complete, excluding from the set of regressors any imputed variable (like income). However, in subsequent iterations the model uses again cases where the dependent variable is complete, but now including the imputed values of the relevant missing regressors (income, in the example above). STATA does not use any convergence measure to determine when convergence is achieved. Hence, a set of 10 iterations was used, and 5 different values were produced. To highlight the fact that non-response may be an interesting outcome in the ECF all imputations are provided as a separate file, along with indicators of whether each value is imputed or if it comes from a complete observation.

References

ALESSIE, R., A. LUSARDI and M. VAN ROOIJ (2011). «Financial literacy and stock market participation», *Journal of Financial Economics*, Volume 101, Issue 2, Pages 449-472.

LITTLE, R. J. A., and D. B. RUBIN (1987). «Statistical analysis with missing data», *Wiley, New York*.

LUSARDI, A., and O. MITCHELL (2011). «Financial literacy around the world: an overview», *Journal of Pension Economics and Finance*, Volume 10 Issue 4, Pages 497-508.

RUBIN, D. B. (1976). «Inference and missing data», *Biometrika*, 63, pp. 581-592

STANGO, V., and J. ZINMAN (2009). «Exponential growth bias and household finance», *Journal of Finance*, Volume 64, Issue 6, Pages 2807-2849.

NUMBER OF QUESTIONS ASKED AND ANSWERED PER SAMPLE INDIVIDUAL, UNWEIGHTED

TABLE 1

	Average	Median	Standard deviation	Minimum	Maximum	Observations
Panel A: Whole sample						
No. of questions asked	115.2	113	14.9	84	191	8,554
No. of questions answered	114.9	112	15.0	81	191	8,554
% of questions answered	99.7	100	1.1	71.1	100	8,554
Panel B: Cases in which only one person was interviewed						
No. of questions asked	111.9	112	6.9	84	127	8,038
No. of questions answered	111.6	112	7.1	81	127	8,038
% of questions answered	99.7	100	1.0	77.0	100	8,038
Panel C: Cases in which two persons were interviewed (a)						
No. of questions asked						
to the sample person	91.1	92	6.3	70	104	515
to the informed person	76.0	76	6.4	52	88	515
No. of questions answered						
by the sample person	90.9	92	6.4	69	104	515
by the informed person	75.4	76	7.4	8	87	515
% of questions answered						
by the sample person	99.8	100	0.6	93.4	100	515
by the informed person	99.2	100	4.8	15.4	100	515

SOURCE: ECF (2016).

a In those 515 cases in which two persons were interviewed, and the informed person agreed to answer (in 1 case no information was given by the informed person).

NUMBER OF COMPLETED INTERVIEWS BY MONTH OF FIELDWORK PERIOD

TABLE 2

Month	No. of interviews	Percent
September	638	7.46
October	2,344	27.40
November	2,331	27.25
December	1,381	16.14
January	812	9.49
February	585	6.84
March	369	4.31
April	69	0.81
May	25	0.29
TOTAL	8,554	100

SOURCE: ECF (2016).

NUMBER OF ATTEMPTED CONTACTS, BY TYPE OF RESPONSE

TABLE 3

	Number of individuals	Average number of contacts	Percentage during the weekend (%)
Valid completed	8,554	3.35	7.31
Refused	6,708	5.91	9.48
Other non-response (a)	763	2.63	8.17
Non-contacted (b)	195	9.78	15.64
Non-eligible (c)	5,030	3.57	8.80
TOTAL	21,250	4.25	8.45

SOURCE: ECF (2016).

a It includes prolonged absence, linguistic barrier, mental incapacity, and discarded after supervision and/or revision.
b It includes absent / not available.
c It includes errors in frame, empty houses, inaccessible and unreachable addresses, or individuals that have moved to an unknown address.

MEASURES OF NON-PARTICIPATION (%)

TABLE 4

	%
Response rate (a)	52.74
Contact rate (b)	98.80
Cooperation rate (c)	53.38
Refusal rate (d)	41.36
Non-eligibility rate (e)	23.67

SOURCE: ECF (2016).

a Computed as (Valid completed) / (Eligible). Eligible includes valid completed, refused, other non-response, and non-contacted.
b Computed as (Contacted) / (Eligible). Contacted includes valid completed, refused, and other non-response.
c Computed as (Valid completed) / (Contacted).
d Computed as (Refused) / (Eligible).
e Computed as (Non-eligible) / (Eligible + Non-eligible). Non-eligible cases include errors in frame, empty houses, inaccessible and unreachable addresses, or individuals that have moved to an unknown address.

PARAMETER ESTIMATES OF A LINEAR REGRESSION MODEL FOR THE PROBABILITY OF ACCEPTANCE TABLE 5

	Coef.	Standard Error	t-ratio
Building condition	0.097	0.033	2.91
Good	0.143	0.041	3.51
In need of some maintenance	0.200	0.059	3.37
Very poor			
Type of area			
High-standing	-0.053	0.054	-0.98
Medium	-0.072	0.067	-1.08
Medium-low	-0.095	0.069	-1.37
Low	-0.058	0.069	-0.84
Size of municipality			
2,000 < inhab = < 10,000	-0.021	0.017	-1.19
10,000 < inhab = < 50,000	-0.029	0.022	-1.32
50,000 < inhab = < 100,000	-0.087	0.029	-2.95
100,000 < inhab = < 500,000	-0.078	0.031	-2.50
500,000 < inhab = < 1,000,000	0.006	0.056	0.10
inhab > 1,000,000	-0.300	0.045	-6.68
Region			
Aragón	0.279	0.074	3.77
Asturias	0.035	0.078	0.45
Balearic Islands	0.073	0.039	1.89
Canary Islands	-0.028	0.106	-0.27
Cantabria	0.067	0.109	0.61
Castile-La Mancha	0.119	0.079	1.51
Castile-León	0.100	0.061	1.63
Catalonia	0.041	0.050	0.80
Valencia	0.076	0.071	1.07
Extremadura	0.098	0.060	1.64
Galicia	0.087	0.085	1.03
Madrid	0.117	0.047	2.47
Murcia	0.127	0.101	1.25
Navarre	0.083	0.088	0.93
Basque Country	0.022	0.066	0.32
La Rioja	0.231	0.059	3.93
Interviewees' characteristics			
Female	-0.029	0.006	-4.46
Age			
30-40	-0.050	0.014	-3.65
40-50	-0.060	0.011	-5.19
50-60	-0.062	0.011	-5.56
60-79	-0.134	0.015	-9.12

PARAMETER ESTIMATES OF A LINEAR REGRESSION MODEL FOR THE PROBABILITY OF ACCEPTANCE (cont'd) TABLE 5

	Coef.	Standard Error	t-ratio
Interviewers' characteristics			
Female	-0.043	0.038	-1.12
Age			
36-45	-0.082	0.059	-1.37
46-55	-0.055	0.066	-0.83
56-65	-0.234	0.079	-2.97
Number of assigned cases	-0.001	0.000	-7.88
Standardised final grade at training test	-0.001	0.037	-0.02
Experience as interviewer			
Years of experience	0.022	0.007	3.43
Years of experience squared	-0.001	0.000	-3.31
Highest level of education			
Upper secondary	0.123	0.061	2.03
Vocational training	0.017	0.061	0.27
University	0.034	0.063	0.54
Type of contract			
Permanent	0.025	0.028	0.90
Number of observations	15,952 of which 8,554 yes (53.6%)		
R^2	0.09		

SOURCE: ECF (2016).

NOTES: The omitted categories are: luxury building, very high-standing neighbourhood, municipalities with 2,000 inhabitants or less, Andalusia, for interviewees' characteristics: male, aged less than 30, and for interviewers' characteristics, male, aged less than or equal to 35, lower secondary education or less, and hired just for the project with a temporary contract. Standard errors are clustered at the interviewer level and adjusted by heteroscedasticity. The sample excludes noneligible cases.

SELECTED INTERVIEWERS' CHARACTERISTICS

TABLE 6

Interviewers' characteristics	Percentage (%)
Female	75.5
Males	24.5
Age	
≤ 35	13.3
36-45	38.8
46-55	30.6
56-65	17.3
Education	
Lower secondary education or less (Inferior a bachillerato)	4.1
Upper secondary education (Bachillerato)	27.5
Vocational training (Formación Profesional)	28.6
Tertiary education (Estudios universitarios)	39.8
Experience as interviewer	
Less than a year	5.1
1 to 5 years	22.5
More than 5 years	72.5
Type of contract	
Permanent	33.7
# of interviewers with at least one valid interview	98

SOURCE: ECF (2016).

REPORTING RATES (%) OF VARIOUS ITEMS, UNWEIGHTED SAMPLE

TABLE 7

	Have item		Value for those having the item				
	Yes	Un-known	Only Point value	Own interval	DK (a)	NA (b)	NP/NF (c)
Monthly payment, loan on main residence	27.99	0.21	89.07	5.55	1.54	3.84	0.29
Loan to value for own house purchase	25.31	0.16	90.43	3.36	5.24	0.97	0.6
Rent main residence	16.31	0.24	93.98	2.29	0.79	2.94	0.72
Household income	100.0	0.0	89.48	0.83	3.45	6.15	0.09
Food expenditure	100.0	0.0	81.19	12.46	5.46	0.71	0.05
Education expenditure	55.12	2.78	98.81	1.38	0.0	0.0	0.0

SOURCE: ECF (2016).

a DK: Do not know.
b NA: No answer.
c NP / NF: Not plausible / Not formulated.

BANCO DE ESPAÑA PUBLICATIONS

OCCASIONAL PAPERS

1401 JOSÉ MARÍA SERENA and EVA VALDEOLIVAS: Integración financiera y modelos de financiación de los bancos globales.

1402 ANTONIO MONTESINOS, JAVIER J. PÉREZ and ROBERTO RAMOS: El empleo de las Administraciones Públicas en España: caracterización y evolución durante la crisis.

1403 SAMUEL HURTADO, PABLO MANZANO, EVA ORTEGA and ALBERTO URTASUN: Update and re-estimation of the Quarterly Model of Banco de España (MTBE).

1404 JUAN CARLOS BERGANZA, IGNACIO HERNANDO and JAVIER VALLÉS: Los desafíos para la política monetaria en las economías avanzadas tras la Gran Recesión.

1405 FERNANDO LÓPEZ VICENTE and JOSÉ MARÍA SERENA GARRALDA: Macroeconomic policy in Brazil: inflation targeting, public debt structure and credit policies.

1406 PABLO HERNÁNDEZ DE COS and DAVID LÓPEZ RODRÍGUEZ: Tax structure and revenue-raising capacity in Spain: A comparative analysis with the UE. (There is a Spanish version of this edition with the same number).

1407 OLYMPIA BOVER, ENRIQUE CORONADO and PILAR VELILLA: The Spanish survey of household finances (EFF): description and methods of the 2011 wave.

1501 MAR DELGADO TÉLLEZ, PABLO HERNÁNDEZ DE COS, SAMUEL HURTADO and JAVIER J. PÉREZ: Extraordinary mechanisms for payment of General Government suppliers in Spain. (There is a Spanish version of this edition with the same number).

1502 JOSÉ MANUEL MONTERO y ANA REGIL: La tasa de actividad en España: resistencia cíclica, determinantes y perspectivas futuras.

1503 MARIO IZQUIERDO and JUAN FRANCISCO JIMENO: Employment, wage and price reactions to the crisis in Spain: Firm-level evidence from the WDN survey.

1504 MARÍA DE LOS LLANOS MATEA: La demanda potencial de vivienda principal.

1601 JESÚS SAURINA and FRANCISCO JAVIER MENCÍA: Macroprudential policy: objectives, instruments and indicators. (There is a Spanish version of this edition with the same number).

1602 LUIS MOLINA, ESTHER LÓPEZ y ENRIQUE ALBEROLA: El posicionamiento exterior de la economía española.

1603 PILAR CUADRADO and ENRIQUE MORAL-BENITO: Potential growth of the Spanish economy. (There is a Spanish version of this edition with the same number).

1604 HENRIQUE S. BASSO and JAMES COSTAIN: Macroprudential theory: advances and challenges.

1605 PABLO HERNÁNDEZ DE COS, AITOR LACUESTA and ENRIQUE MORAL-BENITO: An exploration of real-time revisions of output gap estimates across European countries.

1606 PABLO HERNÁNDEZ DE COS, SAMUEL HURTADO, FRANCISCO MARTÍ and JAVIER J. PÉREZ: Public finances and inflation: the case of Spain.

1607 JAVIER J. PÉREZ, MARIE AOURIRI, MARÍA M. CAMPOS, DMITRIJ CELOV, DOMENICO DEPALO, EVANGELIA PAPAPETROU, JURGA PESLIAKAITĖ, ROBERTO RAMOS and MARTA RODRÍGUEZ-VIVES: The fiscal and macroeconomic effects of government wages and employment reform.

1608 JUAN CARLOS BERGANZA, PEDRO DEL RÍO and FRUCTUOSO BORRALLO: Determinants and implications of low global inflation rates.

1701 PABLO HERNÁNDEZ DE COS, JUAN FRANCISCO JIMENO and ROBERTO RAMOS: The Spanish public pension system: current situation, challenges and reform alternatives. (There is a Spanish version of this edition with the same number).

1702 EDUARDO BANDRÉS, MARÍA DOLORES GADEA-RIVAS and ANA GÓMEZ-LOSCOS: Regional business cycles across Europe.

1703 LUIS J. ÁLVAREZ and ISABEL SÁNCHEZ: A suite of inflation forecasting models.

1704 MARIO IZQUIERDO, JUAN FRANCISCO JIMENO, THEODORA KOSMA, ANA LAMO, STEPHEN MILLARD, TAIRI RÕÕM and ELIANA VIVIANO: Labour market adjustment in Europe during the crisis: microeconomic evidence from the Wage Dynamics Network survey.

1705 ÁNGEL LUIS GÓMEZ and M.ª DEL CARMEN SÁNCHEZ: Indicadores para el seguimiento y previsión de la inversión en construcción.

1706 DANILO LEIVA-LEON: Monitoring the Spanish Economy through the Lenses of Structural Bayesian VARs.

1707 OLYMPIA BOVER, JOSÉ MARÍA CASADO, ESTEBAN GARCÍA-MIRALLES, JOSÉ MARÍA LABEAGA and ROBERTO RAMOS: Microsimulation tools for the evaluation of fiscal policy reforms at the Banco de España.

1708 VICENTE SALAS, LUCIO SAN JUAN and JAVIER VALLÉS: The financial and real performance of non-financial corporations in the euro area: 1999-2015.

1709 ANA ARENCIBIA PAREJA, SAMUEL HURTADO, MERCEDES DE LUIS LÓPEZ and EVA ORTEGA: New version of the Quarterly Model of Banco de España (MTBE).

1801 ANA ARENCIBIA PAREJA, ANA GÓMEZ LOSCOS, MERCEDES DE LUIS LÓPEZ and GABRIEL PÉREZ QUIRÓS: A short-term forecasting model for the Spanish economy: GDP and its demand components.

1802 MIGUEL ALMUNIA, DAVID LÓPEZ-RODRÍGUEZ and ENRIQUE MORAL-BENITO: Evaluating the macro-representativeness of a firm-level database: an application for the Spanish economy.

1803 PABLO HERNÁNDEZ DE COS, DAVID LÓPEZ RODRÍGUEZ and JAVIER J. PÉREZ: The challenges of public deleveraging. (There is a Spanish version of this edition with the same number).

1804 OLYMPIA BOVER, LAURA CRESPO, CARLOS GENTO and ISMAEL MORENO: The Spanish Survey of Household Finances (EFF): description and methods of the 2014 wave.

1805 ENRIQUE MORAL-BENITO: The microeconomic origins of the Spanish boom.

1806 BRINDUSA ANGHEL, HENRIQUE BASSO, OLYMPIA BOVER, JOSÉ MARÍA CASADO, LAURA HOSPIDO, MARIO IZQUIERDO, IVAN A. KATARYNIUK, AITOR LACUESTA, JOSÉ MANUEL MONTERO and ELENA VOZMEDIANO: Income, consumption and wealth inequality in Spain. (There is a Spanish version of this edition with the same number).

1807 MAR DELGADO-TÉLLEZ and JAVIER J. PÉREZ: Institutional and economic determinants of regional public debt in Spain.

1808 CHENXU FU and ENRIQUE MORAL-BENITO: The evolution of Spanish total factor productivity since the Global Financial Crisis.

1809 CONCHA ARTOLA, ALEJANDRO FIORITO, MARÍA GIL, JAVIER J. PÉREZ, ALBERTO URTASUN and DIEGO VILA: Monitoring the Spanish economy from a regional perspective: main elements of analysis.

1810 DAVID LÓPEZ-RODRÍGUEZ and CRISTINA GARCÍA CIRIA: Estructura impositiva de España en el contexto de la Unión Europea.

1811 JORGE MARTÍNEZ: Previsión de la carga de intereses de las Administraciones Públicas.

1901 CARLOS CONESA: Bitcoin: a solution for payment systems or a solution in search of a problem? (There is a Spanish version of this edition with the same number).

1902 AITOR LACUESTA, MARIO IZQUIERDO and SERGIO PUENTE: An analysis of the impact of the rise in the national minimum wage in 2017 on the probability of job loss. (There is a Spanish version of this edition with the same number).

1903 EDUARDO GUTIÉRREZ CHACÓN and CÉSAR MARTÍN MACHUCA: Exporting Spanish firms. Stylized facts and trends.

1904 MARÍA GIL, DANILO LEIVA-LEON, JAVIER J. PÉREZ and ALBERTO URTASUN: An application of dynamic factor models to nowcast regional economic activity in Spain.

1905 JUAN LUIS VEGA (COORD.): Brexit: current situation and outlook.

1906 JORGE E. GALÁN: Measuring credit-to-GDP gaps. The Hodrick-Prescott filter revisited.

1907 VÍCTOR GONZÁLEZ-DÍEZ and ENRIQUE MORAL-BENITO: El proceso de cambio estructural de la economía española desde una perspectiva histórica.

1908 PANA ALVES, DANIEL DEJUÁN and LAURENT MAURIN: Can survey-based information help assess investment gaps in the EU?

1909 OLYMPIA BOVER, LAURA HOSPIDO and ERNESTO VILLANUEVA: The Survey of Financial Competences (ECF): description and methods of the 2016 wave.

BANCO DE **ESPAÑA**
Eurosistema

Unidad de Servicios Auxiliares
Alcalá, 48 - 28014 Madrid
E-mail: publicaciones@bde.es
www.bde.es