BANCO DE ESPAÑA
Eurosistema

DG Economics, Statistics, and Research

21.05.2021

The Survey of Financial Competences (ECF) 2016
User Guide

Microeconomic Studies Division

SUMMARY This document describes the files containing the data from the Survey of Financial Competences (ECF). It also explains how one may proceed about using (i) replicate weights that are made available to take into account sample stratification and clustering and (ii) multiple imputations that are provided to correct for item-non-response in selected variables. A complete description of the ECF and its methods is provided in Bover et al. (2019).

INDEX

1 Data files

1.1 Core data

All the data files are provided in Stata[1] and csv format. The delimiter used in the csv files is a semicolon (;) and the decimal separator is a dot (.).

The files containing the ECF data consist of the following: (i) the main dataset (ecf_2016_e.type[2]), (ii) a separate data set (ecf_2016_imp_e.type[3]) with 5 imputed values of 6 selected variables (i0300_X: monthly rent payment; i0700_X: monthly mortgage payment; i0900_X: loan to value ratio; j0500_X: household food expenditure; j0700_X: household education expenditure; j1300_X: gross annual household income) as well as shadow values indicating whether the particular value was imputed (fi0300, fi0700, fi0900, fj0500, fj0700, fj1300), where *type*=dta, csv.

In *type*=csv, Stata programs for labelling the ECF variables of the two data sets mentioned above are called labels_ecf_2016_e.csv[4] and labels_ecf _2016_imp_e.csv[5]. These files may be used to label variables of the datasets mentioned above.

The individual identifier variable common to all datasets is: ID. Note that the sample unit is the individual.

1.2 Replicate weights

We provide replicate weights to enable users taking into account sampling design features in the estimation of the sampling variances (see below some comments about the use of replicate weights for the calculation of variances).

Replicate weights are stored in the file replicate_weights.*type*. The file contains 1000 replicate cross-section weights (wt3r_*i*, *i*=1,…, 1000) and 1000 multiplicity factors (ntimesr_*i*, *i*=1,…, 1000)[6]. A description of the cross-sectional weights is given in Section 3.

1.3 Main results: tables

The following files are also available:

(i) File containing tables with the main results (pdf file). A first version of those tables based on preliminary imputations was published in the *Main Results of the Survey of Financial Competences*[7].

(ii) Definitions of the variables reported in the tables as Stata commands (Word file).

[1] Stata 12. Stata 11 can also read Stata 12 data sets.
[2] ecf_2016_e.type refers to the English version and ecf_2016.type to the Spanish one.
[3] ecf_2016_imp_e.type refers to the English version and ecf_2016_imp.type to the Spanish one.
[4] labels_ecf_2016_e.csv refers to the English labels and labels_ecf_2016.csv to the Spanish one.
[5] labels_ecf_2016_imp_e.csv refers to the English labels and labels_ecf_2016_imp.csv to the Spanish one.
[6] The multiplicity factor indicates the number of times the observation has been selected in the resampling.
[7] Both Spanish and English versions published by the Banco de España in May 2018. Results in Spanish are available here: https://www.bde.es/f/webbde/SES/estadis/otras_estadis/2016/ECF2016.pdf.
For an English version, see https://www.bde.es/f/webbde/SES/estadis/otras_estadis/2016/ECF2016-en.pdf

2 Variables

The ECF was collected using a computer assisted personal interview (CAPI). A paper version of the CAPI questionnaire is provided on the web site (both in the original Spanish wording and in English). Some variables included in the questionnaire are not provided for anonymity reasons: month and day of birth for the sample member (a0500 and a0600) or month and day of birth of the informed person (g0200 and g0300). Literals are not provided either. Finally, some variables are recoded into broader aggregates to preserve confidentiality, like the detailed country of birth for sample members born outside Spain or the province of birth of those born within Spain (a0300 and a0200)[8] .

2.1 Naming of the questionnaire variables in the Stata files

The questionnaire variables in the data have been named according to some common patterns that should help in identifying the corresponding question.

The variable lnnnn refers to question number nnnn in section l (sections are identified by consecutive letters from a to j).

Questions are numbered using four digits. The first two digits indicate the position of the question in the section, while the latter two being used for sequence of questions encompassed in the same concept. For example, the questions about whether the individual has heard of, holds or has recently acquired a financial product in a list of ten items share the first three positions lnnxx but differ in the last two, as the same question is asked about 10 different assets.

Examples:

The variable b0100 refers to question number 1, section B.

The variable b0401 refers to question number 4 in section B. That question is asked about 10 different financial assets and b0401 is asked about the first financial product (a mortgage).

The variable d1201 refers to the question number 12 in Section D, and within the concept, there are 11 statements, that the respondent must answer using the same scale. The variable d1201 is the first of such statements.

2.2 Variables from questions with multiple answers

For questions with multiple answers we generate variables with a pattern equivalent to the previous one but adding after the number of the question a letter, i.e, notation is (lnnnnl).

Variables lnnnnl (where the letter l in the sixth position reflects consecutive letters from a to q) correspond to questions where as many dummy variables are generated as the number of possible responses.

[8] The public version of the variable "a0300" distinguishes immigrants born in Europe from those coming from the rest of the world while that of a0200 distinguishes region (CCAA) of birth.

Examples:

Variable b1000c refers to the question number 10 in Section B, vehicles used for saving during the last 12 months, and third possible answer (whether the individual has been saving in a term accounts). The variable takes value 0 if the individual has not chosen an option and 1 if he or she has chosen it -we deal with cases of non-response below.

Variable b1202a refers to the second part of question number 12 in section B, first derived question a (first possible form of financial discrepancies with a financial institution). When the b1202a takes value 1 the respondent has answered having had that specific problem in the last five years, and zero otherwise.

2.3 Additional variables in the dataset

The data includes some constructed variables that are specified below.

- The variable age reflects the age in years of the respondent sample member at the moment of the interview. It is constructed by subtracting from the year and month of the interview the year and month of birth (and day of birth when the month of the interview coincides with the month of birth). Individuals who do not report their exact year or month of birth are asked their approximate age in years. The reported variable "age" is provided for each individual, and it is the derived variable from exact year and month of birth -for respondents who gave that information- and the direct response of individuals who chose not to answer exact year and month of birth. The variable was used to adjust sample weights by non-response.
- The variable ccaaf refers to the autonomous community code, where the sample person was interviewed.
- The variables tmp_e0401 and tmp_h0401 were generated by the computer application used during personal interviews. These variables are useful tools to capture reading comprehension in questions e040$ and h040$ (see questionnaire). They measure the time required by the interviewee to read and understand a written text unrelated to financial matters.

2.4 Codes for non-response

The ECF contains a set of codes to reflect the type of non-response to each question.

These values are -3, -4, -5, -96, -97, -98, -99. Their meanings are as follows:

-3: The individual cannot read.
-4: Value considered unreliable.
-5: The individual does not answer a question in the questionnaire due to CAPI or interviewer error.
-96: Variable set to missing for anonymisation (like, country of birth for individuals born outside Spain).
-97: The answer is ´Don't know´.
-98: True missing, derived from the answer given by the household on a previous variable in the questionnaire or to a lack of response to a previous question that acts as a filter.
-99: The answer is ´No Answer´.

3 Weights

We provide one set of cross-sectional weights (*weight*) to compensate for unequal probability of the individual being selected into the sample given the geographical stratification, and differential unit non-response. The sum of weights over all individuals in the sample is an estimate of the total number of individuals in the population at October 2016Q4 (i.e. the weights reported are the inverse of the probability that an individual is in the sample).

Taking into account weights is crucial in obtaining population totals and means from the ECF data. However, there is some controversy on when weights should be used in regressions [Deaton (1997, Chapter 2) and Cameron and Trivedi (2005, Chapter 24) provide a very useful discussion on these issues]. Each user has to evaluate the situation given the objectives of the analysis at hand.

Note that when analysing small fractions of the sample, care should be taken in applying weights which have been constructed for the whole sample.

4 Standard error calculations

Samples designed for surveys rarely consist in simple random sampling from the population. They usually involve some (i) stratification and/or (ii) clustering. To calculate the sampling variance of estimates of interest one needs to take into account these characteristics of the sample design. Stratification may increase the precision of estimates over simple random sampling if, for example, means are different across strata. Some clustering (i.e. sampling first clusters or primary sampling units – *secciones censales* – and then choosing individuals from within each cluster) is usual sampling practice in order to reduce costs but it may diminish precision if individuals' characteristics are similar within clusters. Therefore, the use of standard random sample formulas for evaluating the sampling variance may be misleading.

For simple sample designs and simple statistics appropriate variance formulas can be derived using Taylor approximations. Alternatively, bootstrap is a more computer intensive method widely used [first introduced in Efron (1979); see Horowitz (2001)]. Bootstrap samples repeatedly from the original sample with replacement. The drawing of these repeated samples is done taking into account the sample design. At each resampling the statistic of interest is evaluated and stored. The variability of these resampling statistics is used as a measure of the variance of the original sample statistic.

However, taking stratification and clustering sampling features into account, either analytically or by bootstrapping, requires the availability of stratum and cluster indicators. Generally, Statistical Offices or survey agencies do not make them available for confidentiality reasons.

Alternatively, to enable more accurate variance estimates with the ECF data without disclosing stratum or cluster information we provide 1000 replicate weights. This number of replicates is regarded as sufficient to estimate the tails of the distribution. For variance estimation a smaller number would be needed.

With a set of replicate weights, the variance can be estimated from repeated estimation of the statistic of interest for each of the 1000 replicate weights. This is an alternative to 1000 bootstrap resampling estimates using stratum and cluster indicators (and a unique weight).

Below we include some Stata code as an example on how one could proceed to estimating the standard error of the mean of a particular variable (the average proportion of correct answers to the question on interest rate compounding).

```
/* STANDARD ERRORS USING REPLICATE WEIGHTS HERE, FOR EXAMPLE, FOR THE MEAN */

* A. OBTAINING THE STANDARD ERROR OF A VARIABLE, HERE, FRACTION OF INDIVIDUALS WITH CORRECT ANSWER IN INTEREST RATE COMPOUNDING
* To compute the statistic, we estimate a linear regression of the variable on a constant

/* Adding information on replicate weights data*/
use "C:\ecf_2016.dta", clear
sort ID
merge 1:1 ID using "C:\ECF_replicate_weights.dta"
tab _merge
```

```
save "C:\ECFreplw.dta", replace

* Obtaining the mean using population weights
gen intcomp_correct=(e0900==1)
reg intcomp_correct [pw=weight]

* Obtaining the standard error with replicate weights
* First bootstrap sample and its weighted mean
reg intcomp_correct [pw=FACTOR_1]
gen Mintcomp_correct=_b[_cons]
list Mintcomp_correct in 1/2
drop if _n>1
keep Mintcomp_correct
save "C:\loop1.dta", replace
clear
* Reps-1000 bootstrap samples and their weighted means
set output error
forvalues s=2/1000 {
        use "C:\ECFreplw.dta", clear
        gen intcomp_correct=(e0900==1)
        reg intcomp_correct [pw=FACTOR_`s']
        gen Mintcomp_correct=_b[_cons]
        list Mintcomp_correct in 1/2
        drop if _n>1
        keep Mintcomp_correct
        append using "C:\loop1.dta"
        save "C:\loop1.dta", replace
        drop _all
}
set output proc
use "C:\loop1.dta"
*The sum command will provide the sampling standard error of the mean
sum
clear
```

For the most common estimation commands, Stata users can also make use of the svy command to estimate the variance using replicate weights. First, we need to merge our dataset with the replicate weights file in a similar way as that indicated in the example above for constructing the dataset ECFreplw, and to specify the replicate weights using the svyset command.

Below we show an alternative way of estimating the variance for the sample weighted mean of the variable analyzed in the example above -i.e., the fraction of individuals answering correctly the question on interest rate compounding.

```
/* STANDARD ERRORS USING REPLICATE WEIGHTS AND THE STATA SVY COMMAND. SAME EXAMPLE */
```

```
*A.- Statistic of interest: the mean of correct responses to the question on interest rate compounding
* To compute the statistic, we estimate a linear regression of the variable on a constant

* We first merge our data with the replicate weights file as indicated in the example above
import delimited "C:\ECF_replicate_weights_csv.csv", delimit(";") clear
rename id ID
sort ID
save "C:\ECF_replicate_weights_csv1.dta", replace

import delimited "C:\ecf_2016.csv", delimit(";") clear
rename id ID
sort ID

merge 1:1 ID using "C:\ECF_replicate_weights_csv1"
tab _merge
drop _merge
save "C:\ECFreplw.dta", replace

gen intcomp_correct=(e0900==1)
svyset [pweight=weight]
svy: reg intcomp_correct

* B. Obtaining the standard error of the computed mean.
use "C:\ECFreplw.dta", clear
gen intcomp_correct=(e0900==1)
svyset [pweight=weight], bsrweight(factor_*) vce(bootstrap)
svy: reg intcomp_correct
```

5 Imputation

5.1 The imputed dataset

For a small group of six questions in the ECF measuring household income or expenses that the respondent replied "Do not know" (DK), "No answer" (NA), unreliable values were given or were not asked because of a CAPI or interviewer error, we provide a set of imputed variables. The variables are: i0300_X: monthly rent payment; i0700_X: monthly mortgage payment; i0900_X: loan to value ratio; j0500_X: household food expenditure; j0700_X: household education expenditure; and j1300_X: gross annual household income.

The use of imputed values enables the analysis of the data with complete-data methods. However, the user is free to ignore the imputations we provide and obtain his/her own or work with explicit probability models for non-response (imputed values are identifiable through the corresponding shadow variable, as described above). For an introduction to the reasons for imputation and the choice of the imputation method used in the ECF see Bover et al (2019).

For each missing value (i.e. NA/DK answer) we provide five imputed values. These imputations are stored as five distinct version of the variable (five 'implicates'). One distinct advantage of using multiple imputations (MI) is to be able to assess the uncertainty associated with the imputation process [see Rubin (1987)].

For each variable, there can be two cases. Either the respondent answer the question -and in this case, the original response is the same in the 5 variables- or the respondent did not answer -in which case, the value of the response may vary across the different imputations -see Bover et al (2019).

The convention for naming those variables follows the structure of the questionnaire Innnn_X, where X is a number going from 1 to 5. The latter number signals the replicate number.

The variable Innnn_X will generally take numerical responses, with three exceptions. If the question does not apply, the value of Innnn_X will be -98. For example, individuals who own the house they live in should not provide any answer to a question on the monthly amount paid as rent. The second exception is when the individual did not answer a preceding question (for example, if he or she owns her accommodation), in which case, the variable will also take -98.

For example, the imputed variable j1300_1 denotes the 13th question in the J section (household income). The suffix 1 indicates that it is the first of the five imputations. To obtain a reliable indicator of the income bracket of the respondent one must combine the five values of j1300_X, as described in Section 5.2.

Along with those variables we provide a "flag" variable for each variable Innnn_X -i.e., a variable indicating whether the value collects the original response by the interviewer -in which case the five values of each Innnn_X will be the same or, alternatively, if the respondent did not answer to the original variable, in which case the five values Innn_X may be different. The name of this flag variable is flnnnn_X, and takes values -98, 0 and 1.

Their meanings are as follows:

-98: That question is not applicable.

0: That question has been directly replied by the interviewee.

1: The value for variable Innnn has been imputed.

5.2 Using the imputed data

To make inferences from the five multiply imputed values one has (1) first to analyse each of the five values by complete-data methods and (2) then combine the results.

Suppose the interest lies in a point estimate of some parameter Q (e.g. mean, median, regression parameter) and that for each of the five imputed values we have obtained an estimate of Q (using standard complete-data methods), denoted $\hat{Q}_i$. The MI point estimate of Q, $\bar{Q}$, is the average of the five complete data estimates

$$\bar{Q} = 1/5 \sum \hat{Q}_i$$

The variance associated with this estimate $\bar{Q}$ has two components:

(i) the within imputation sampling variance W which is the average of the five complete-data variance estimates $(\hat{V}_i)$:

$$W = 1/5 \sum \hat{V}_i$$

(ii) the between imputations variance B which reflects the variability due to imputation uncertainty and is the variance of the complete data point estimates:

$$B = 1/4 \sum (\hat{Q}_i - \bar{Q})^2$$

The total variance for $\bar{Q}$ is given by:

$$T = W + (6 / 5)B$$

In practice, to obtain MI estimates of the type just described, the user may find useful some of the following alternatives:

(i) Stata users may find helpful to download and use the procedures described in Carlin et al. (2003, 2008) for manipulating and analysing MI datasets.

(ii) Stata users can also make use of the mi import, mi estimate or mim commands provided by Stata after version 11 for estimating and analysing descriptive statistics using a unique dataset that pools together the five imputed datasets (see example below).

(iii) Finally, for general modelling outcomes, the user has to perform the analysis five times and combine them following the formulae above. To help see the simplicity of combining the results from the five datasets we include below few lines of Stata code that would provide the combined results (MI point estimate and its standard error) from inputting the five point estimates and five standard errors.

Usually it may suffice to do the exploratory analysis with one or two of the MI datasets and only use all of the five datasets for final results.

```
* STATISTIC OF INTEREST: COMPUTING THE MEDIAN AND STANDARD ERROR OF AN IMPUTED VARIABLE USING A SINGLE IMPLICATE. EXAMPLE: WE USE THE MONTHLY AMOUNT PAID AS MORTGAGE (i0700_)

* Explaining how to generate the input file
import delimited "C:\ecf_2016_imp.csv", clear delim(";")
rename id ID
sort ID
save "C:\ecf_2016_imp1.dta", replace

import delimited " C:\ECF_replicate_weights_csv.csv", clear delim(";")
rename id ID
sort ID
save " C:\ECF_replicate_weights.dta", replace

import delimited "C:\ecf_2016.csv", clear delim(";")
rename id ID
sort ID
merge 1:1 ID using "C:\ecf_2016_imp1.dta"
keep ID weight i0700_1 i0700_2 i0700_3 i0700_4 i0700_5
sort ID
merge 1:1 ID using "C:\ECF_replicate_weights.dta"
keep ID weight i0700_1 i0700_2 i0700_3 i0700_4 i0700_5 FACTOR_*
save "C:\ecfimp_weights.dta", replace

* First implicate
qreg i0700_1 if i0700_1!=-98 [pw=weight]
gen Mmortgage=_b[_cons]
keep Mmortgage
gen imp=1
keep imp Mmortgage
drop if _n>1
sort imp
save "C:\input_imp1.dta", replace

* First bootstrap sample and its weighted median
use "C:\ecfimp_weights.dta", clear
qreg i0700_1 if i0700_1!=-98 [pw= factor_1]
gen rMortgage=_b[_cons]
list rMortgage in 1/2
drop if _n>1
keep rMortgage
save "C:\loop1.dta", replace
clear

* Reps-1000 bootstrap samples and their weighted medians
set output error
```

```
forvalues s=2/1000 {
        use "C:\ecfimp_weights.dta", clear
        qreg i0700_1 if i0700_1!=-98  [pw=factor_`s']
        gen rMortgage=_b[_cons]
        list rMortgage in 1/2
        drop if _n>1
        keep rMortgage
        append using "C:\loop1.dta"
        save "C:\loop1.dta", replace
        drop _all
}

set output proc
use "C:\loop1.dta"

*The sum command will provide the sampling standard error of the median of the first implicate of the mortgage instalment (i0700_1)
sum
gen seMortgage=r(sd)
drop if _n>1
gen imp=1
keep seMortgage imp
sort imp
merge imp using "C:\input_imp1.dta"
save "C:\input.dta", replace
clear

* The same procedure must be repeated for the five implicates, appending the resulting estimates of the median and its standard error in the dataset as "input.dta"
```

```
*OVERALL ESTIMATES OF THE STANDARD ERROR OF THE MEDIAN OF i0700 (option (iii) above);
use c:\input.dta
*the file input.dta should contain five observations and two variables which are the point estimate (called here Mmortgage) and the standard error (called here seMortgage) for each of the five datasets
gen ni=5
set type double
* within component is the average of SE from each implicate
gen varmean=seMortgage*seMortgage
egen w=mean(varmean)
*qbar denotes the overall point estimate
egen qbar=mean(Mmortgage)
* between component is the variance of the point estimates
gen dev=(Mmortgage-qbar)*(Mmortgage-qbar)
egen be=sum(dev)
replace be=be*(1/(ni-1))
*totvar denotes the overall variance (within and between component)
```

```
gen totvar=w+(1+(1/ni))*be
*sqrttotvar denotes the overall standard error
gen sqrttotvar=sqrt(totvar)
format qbar totvar sqrttotvar %12.1f
list
```

---

*USING MI STATA COMMANDS (option (ii) above)

As mentioned above, an alternative way of combining results is to work with a unique dataset that pools together the five imputed datasets. To make use of the mi command in Stata, the unique dataset must contain an additional variable indicating from which multiply imputed data set the observation comes from (for example, an indicator called *mdataset* taking value 1 if the observation comes from the first imputation, value 2 if it comes from the second and so on). In addition, Stata also requires the inclusion of an additional dataset containing only original data, whose observations are marked with value 0 in the indicator mentioned above.

However, if Stata users do not wish to work with the original data, they can duplicate the first dataset provided by the ECF and set the indicator to 0, as if these were the original data. Before making use of the mi estimate command, we need to import the unique ECF data multiply imputed as follows: use ecf; mi import flong, m(mdataset) id(ID), assuming *ecf* is the name of this pooled dataset.[9]

We include below an example estimating the sample weighted median of a variable using the mi command in Stata.

```
* COMBINED ESTIMATES USING THE MI COMMAND IN STATA
 * A PARTICULAR CASE: COMPUTING THE MEDIAN OF AN IMPUTED VARIABLE USING STATA.
WE USE THE MONTHLY AMOUNT PAID AS MORTGAGE (i0700)

* A. Reading the imputed dataset

import delimited "C:\ecf_2016_imp.csv", clear delim(";")
rename id ID
sort ID
keep ID weight i0700_1 i0700_2 i0700_3 i0700_4 i0700_5
save "C:\ecf_2016_imp1.dta", replace


** Piling up the 5 implicates for each variable
reshape long i0700_, i(ID) j(mdataset)


save "C:\ecf15", replace


*B. Generating the 0 dataset and append it to the pooled dataset

use "C:\ecf15", clear
keep if mdataset==1
replace mdataset=0 if mdataset==1
```

[9] It is convenient to include the option “esampvaryok” in all mi estimate commands, in order to avoid an error message that appears when the estimation sample varies across implicates, as documented in Stata. Another useful option of the mi estimate command is “post” (mi estimate, esampvaryok post:...), mainly when we wish to use postestimation commands in Stata.

```
append using "C:\ecf15"
save "C:\ecf05", replace
export delimited using "C:\ecf05.csv", delimiter(";") replace

* C. Using the mi Stata command (option (ii) above)[10]
mi import flong, m(mdataset) id(ID)
mi estimate, esampvaryok post: qreg i0700_ if i0700_!=-98 [pw=weight]
```

To use replicate weights in a unique dataset using the mi estimate command, we need to make use of another option of the mi estimate command, called "vceok" (this is not documented by Stata). An example of the commands needed to implement this alternative for estimating W directly is as follows.

```
* COMBINING STANDARD ERRORS USING REPLICATE WEIGHTS AND THE MI COMMAND
* A PARTICULAR CASE: COMPUTING THE MEAN OF AN IMPUTED VARIABLE USING STATA. *
WE USE THE IMPUTED MONTHLY AMOUNT PAID AS MORTGAGE (i0700_)

* A. GENERATE A SINGLE DATASET
*For the dataset with value 0 in the multiply imputed dataset indicator, we use the file ecf05

*B. COMBINING RESULTS USING MI COMMAND;

import delimited "C:\ecf05.csv", clear delim(";")
rename id ID
sort ID
save "C:\ecf05.dta", replace

import delimited "C:\ECF_replicate_weights_csv.csv", clear delim(";")
rename id ID
sort ID
merge 1:m ID using "C:\ecf05.dta"
save "C:\ecf05replw.dta", replace
export delimited using "C:\ecf05replw.dta", delimiter(";") replace

mi import flong, m(mdataset) id(ID)
mi svyset [pweight=weight], bsrweight(factor_*) vce(bootstrap)
mi estimate, esampvaryok post vceok: svy: reg i0700 if i0700!=-98
```

The dataset ecf05 is the result of appending the five datasets with each of the values of the imputed variables having previously generated the imputed dataset indicator, together with a duplicated dataset containing the value 0 in the imputed dataset indicator, as illustrated in one of the examples above.

[10] The same example but for the mean instead of for the median would replace the line of command "mi estimate, esampvaryok post: qreg i0700_ if i0700_!=-98 [pw=weight]" with "mi estimate, esampvaryok post: reg i0700_ if i0700_!=-98 [pw=weight]".

REFERENCES

BOVER, O. HOSPIDO, L. and VILLANUEVA, E. (2019) 'The Survey of Financial Competences (ECF): Description and methods of the 2016 wave', Occasional paper N.º 1909, *Banco de España*.

CAMERON, A. C., and P. K. TRIVEDI (2005). *Microeconometrics: Methods and Applications,* Cambridge University Press.

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

CARLIN, J. B., J.C. GALATI, and P. ROYSTON (2008). 'A new framework for managing and analyzing multiply imputed data in Stata', *The Stata Journal, 8*, pp. 49-67.

EFRON, B. (1979). 'Bootstrap methods: another look at the jackknife', *Annals of Statistics,* 7, pp. 1-26.

DEATON, A. (1997). *The Analysis of Household Surveys,* The World Bank, The John Hopkins University Press.

HOROWITZ, J. L. (2001). 'The Bootstrap', in *Handbook of Econometrics, Volume 5,* edited by J. J. Heckman and E. Leamer, Elsevier Science.

RUBIN D. B. (1987). *Multiple Imputation for Nonresponse in Surveys,* Wiley.