BANCO DE ESPAÑA
Eurosistema

DG Economics, Statistics, and Research

02.03.2021

The Spanish Survey of Household Finances (EFF) 2017
User Guide

Microeconomic Studies Division

SUMMARY This document describes the files containing the data from the 2017 Spanish Survey of Household Finances (EFF). It also explains how one may proceed about using these files regarding: (i) linking EFF2014-2017 panel households and their members, (ii) multiple imputations that are provided to correct for item-non-response, and (iii) replicate weights that are made available to take into account sample stratification and clustering. A complete description of the 2017 wave and its methods is provided in Barceló et al. (2020).

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 EFF2017 data consist of the following: (i) five imputed data sets, (ii) data set with the shadow variables.

Missing data in the survey have been imputed five times using a multiple imputation procedure. The corresponding data are stored into five separate files: effe_2017_imp1_*type*.zip, effe_2017_imp2_*type*.zip, effe_2017_imp3_*type*.zip, effe_2017_imp4_*type*.zip, and effe_2017_imp5_*type*.zip[2] (*type*=dta, csv).

Each effe_2017_imp*i*_*type*.zip (*i*=1, 2, 3, 4 and 5) file contains the following:
other_ sections_2017_imp*i*.*type*: all sections of the questionnaire except section 6.
section6_2017_imp*i*.*type*: section 6[3].

In *type*=csv, Stata programs for labelling the EFF variables of the two data sets mentioned above are also included in effe_2017_imp1_csv.zip, called labels_other_sections_2017.do _and labels_section6_2017.do, respectively.[4] These programs may be used to label variables of the five imputed datasets mentioned above.

There is also a file, common to all five imputations, containing shadow values of the original variables (sombra_2017.*type*). The purpose of this file is to provide as much information as possible about the original state of the variables. Each variable in the survey has a shadow variable that reflects the information content of the primary variable (see below sub-section 3.4 for more details).

The household identifier variable common to all datasets is: h_2017. Note that the sample unit is the household.

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).

Three files are available: (i) replicate_weights_2017.*type* contains 1000 replicate cross-section weights (wt3r_*j*, *j*=1,…, 1000) and 1000 multiplicity factors (ntimesr_*j*, *j*=1,…, 1000)[5], (ii) replicate_pan1weights_2017.*type* contains 1000 replicate 2014 longitudinal weights (wtpan1r_*j,* *j*=1,…, 1000) and 1000 multiplicity factors (ntimespan1r_*j,* *j*=1,…,1000), (iii) replicate_pan2weights_2017.*type* contains 1000 replicate 2017 longitudinal weights (wtpan2r_*j,*

[1] Stata 12. Stata 11 can also read Stata 12 data sets.
[2] The data files with the variable labels in Spanish (available from our web site in Spanish) are called eff_2017_imp1_*type*.zip,…, eff_2017_imp5_*type*.zip.
[3] These are named otras_secciones_2017_imp*i*.*type* and seccion6_2017_imp*i*.*type* in the Spanish version.
[4] The corresponding programs in the Spanish version are called etiquetas_otras_secciones_2017.do and etiquetas_seccion6_2017.do, and they are provided in eff_2017_imp1_csv.zip.
[5] The multiplicity factor indicates the number of times the observation has been selected in the resampling.

i=1,…, 1000) and 1000 multiplicity factors (ntimespan2r_*i,* *i*=1,…, 1000). A description of the cross-sectional weights and the two sets of panel weights is given below.

1.3 Main results: tables and data used

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 *Analytical Article*[6].

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

(iii) Data files with the above constructed variables (5 of them, one for each imputed data set).

[6] Both Spanish and English versions published by the Banco de España in December 2019.

2 Linking EFF2014-2017 panel observations

The EFF2014 household identifier variable (h_2014) is provided in order to allow the linking of households that participated in both the 2014 and 2017 waves. However, care should be taken in interpreting a panel household as the same household in the two waves since its composition may have changed substantially.

The variable hogarpanel is an indicator that takes the value 1 for households that participated in both waves and zero otherwise.

To identify individual members of a panel household that were part of the household in both waves the variable pan_x, (x=1,…, 9) should be used. This variable takes the following values:

== 0: member number x in 2017 was not a member of the household in the previous wave

== y (y=1,…, 9): member number x in 2017 wave is the same person as member number y in 2014 wave

== missing: not a panel household

To facilitate converting EFF household files into individual member files in order, for example, to link individuals instead of households we provide below some Stata code.

```
-------------------------------------------------------------------------
* FROM HOUSEHOLD TO INDIVIDUAL MEMBER FILES;

use C:\effe.dta;
keep p1  h_2017  p1_1_*  p1_2b_*  p6_32_*_*;
reshape long p1_1_  p1_2b_  p6_32_@_1  p6_32_@_2  p6_32_@_3, i(h_2017)  j(eff17_miem);
rensfix _ ;[7]
sort h_2017 eff17_miem;
by h_2017: drop if _n>p1;
-------------------------------------------------------------------------
```

[7] "rensfix _" removes in this case suffix _ at the end of variable names but this is not essential to the individual members linking described. Contrary to "renpfix" which is a Stata command, "rensfix" is a user written addition (by Jenkins and Cox, 2001) from Stata Technical Bulletin (STB-59).

3 Variables

The EFF 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). Month and place of birth variables (p1_2a_*i*, p1_2c_*i*, p1_2d1_*i*, p1_6_*i*, p1_6a_*i*, p1_6b_*i*, *i*=1,…, 9) that appear in the paper questionnaire are not provided for anonymity reasons. For the same reasons, age has been top coded (and year of birth has been bottom coded accordingly). [8] There are some mop-up questions which are only used for imputation and are not provided. In particular, these are: p2_12_0, p2_18_0, p2_9_0, p2_39_0, p2_43_0, p2_50_0, p2_55_0, p2_61_0, p2_55_0b_*i*, p2_61_0b_*i*, *i*=1,…,3, p3_6_0, p3_11_0, p6_49_0_*i*, *i*=1,…,9. Finally, p7_13a is not in the data because no specific answers were recorded.

3.1 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.

These patterns are the following:

(i) The variable ps_nn refers to question number nn in section s.

(ii) The variable ps_nn_m refers to question number nn in section s. Position m appears when question ps_nn is asked several times. For example, when the same question is asked to each household member.

(iii) The variable ps_nn_m_r refers to question number nn in section s. The letter m has the same meaning as before and position r appears when the question ps_nn_m is asked several times. For example, when details are asked on the characteristics of each self-employed job for each household member.

Examples:

The variable p2_5 refers to question number 5, section 2.

The variable p2_52_2_1 refers to question number 52, section 2, first loan for second property.

The variable p6_3_2 refers to question number 3, section 6, for the second household member.

The variable p6_13_4_2 refers to question number 13, section 6, second paid-employment job of the fourth household member.

[8] Other variable bottom coded for anonymity reasons is p2_29 (the year of start living in the house or flat).

3.2 Variables from questions with multiple answers[9]

For these questions we generate variables with a pattern equivalent to the previous one but adding after the number of the question the codes cX, sX or zX (ps_nncX_m_r, ps_nnsX_m_r or ps_nnzX_m_r).

The use of these codes is determined as follows:

(i) Variables ps_nncX_m_r correspond to questions where as many dummy variables are generated as alternative answers can be given by the respondent.

(ii) Variables ps_nnsX_m_r correspond to questions where it is assumed that each respondent will answer no more than five options. This way, the variables created for each question of this kind are at most five (ending in s1, s2, s3, s4 and s5 in the Stata file). There is no ordering of the answers when more than one option is chosen by the respondent.

(iii) Variables ps_nnzX_m_r correspond to questions where as many variables are generated as the number of future scenarios or events are contemplated in questions.

Examples:

Variable p6_1c2_1 refers to question number 1, section 6, current labour status of the first household member and second possible answer ("self-employed"). This variable can take two values: 0 and 1 (indicating no and yes, respectively).

Variable p6_1c3_1 refers to question number 1, section 6, current labour status of the first household member and third possible answer ("unemployed"). This variable can take two values: 0 and 1.

Variable p2_42s1_4 refers to question number 42.4, section 2, for the first answer of the household (question 2.42.4 of the questionnaire). This variable can take the values 1, 2, 3, 4, 5, 6, 7, 97. Note that for properties number 1, 2, and 3 p2_42 is not a multiple choice question.

Variable p2_42s2_4 refers to question number 42.4, section 2, for the second answer given by the household (if any) (question 2.42.4 of the questionnaire). This variable can take the values 1, 2, 3, 4, 5, 6, 7, 97.

Variable p2_25az2 refers to question number 25a, section 2, for the chances given by the household to the second scenario contemplated (question 2.25a of the questionnaire).

[9] When multiple answers are allowed, the different possible answers are followed by M in the paper questionnaire except for questions that ask respondents about their expectations on some future events (questions 2.25a and 6.60i of the questionnaire). See paragraph (iii) in the main text.

3.3 Constructed total household income variables

Also included in the data are two constructed total household income variables, one corresponding to the whole of 2016 (renthog) and the other to the month (during 2017 or 2018) in which the interview took place (mrenthog).

The latter variable (mrenthog) includes other irregular income earned in the last three months (p6_60f). This source of income cannot be extrapolated to other months of the year. Thus, an annual income at the year of interview cannot be constructed by multiplying mrenthog by 12.

These two constructed income variables are calculated as the sum of labour and non-labour income of all household members. When the household fails to provide a value for one of those components we perform a direct imputation of the total. Given that the income components have also been imputed it is also possible to construct an alternative imputation of total income based on the imputed components, which obviously differs from directly imputed total income.

3.4 Shadow variables

Following the same naming pattern, a series of additional variables have been created (shadow variables) to facilitate the identification of the values that have been imputed. The only difference in the naming of these variables is that they start with “j” instead of “p”.

These variables can take the following values: 0, 1, 2050, 2051, 2052, 2053, and 2055. Their meanings are as follows:

1: complete observation.
0: true missing, derived from the answer given by the household on a previous variable in the questionnaire.
2050: imputed value when the answer is ‘Don’t know’.
2051: imputed value when the answer is ‘No Answer’.
2052: imputed value due to the lack of answer to other preceding variables.
2053: answered by the household but incorrect; value has been imputed.
2055: household does not answer a question contemplated in the questionnaire due to CAPI or interviewer error.

Only those observations with shadow values equal or higher than 2050 are to be imputed.

4 Weights

We provide one set of cross-sectional weights (*facine3*) to compensate for (i) unequal probability of the household being selected into the sample given the oversampling of the wealthy in the EFF and geographical stratification, and (ii) differential unit non-response. In the construction of these weights account is also taken of the household composition and therefore the weight is the same for the household and for any of the household members. The sum of weights over all households in the sample is an estimate of the total number of households in the population at 2017Q4 (i.e. the weights reported are the inverse of the probability that a household is in the sample).

Taking into account weights is crucial in obtaining population totals, means, and shares from the EFF 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 analyzing small fractions of the sample, care should be taken in applying weights which have been constructed for the whole sample.

Additionally to cross-section weights we provide two longitudinal sets of weights that compensate for differential non-response in the EFF2017 of the EFF2014 households. The first set (pesopan_1) is adjusted to conform to the 2014 population and could be used to study transitions from a 2014 representative population to their 2017 situation. The second set (pesopan_2) conforms to the 2017 population and would be of use if interested in studying the 2014 situation of a representative 2017 population. In any case, care should be exercised in trying to infer results that are representative of the 2014 or the 2017 population from the panel subsample. Barceló et al. (2020) details how cross-sectional and longitudinal sample weights are constructed for the EFF.

5 Imputation

Imputations are provided for the 'No Answer' (NA) or 'Don't Know' (DK) replies for all the variables in the survey, except very few variables where the NA/DK category exceeds 60% of the answers to the question or the observations are too few (see Appendix for a list of those variables).

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, see Bover (2004), and for a detailed description of imputation in the EFF, see Barceló (2006). The imputations provided so far are static in the sense that they do not use the information contained in other waves.

For each missing value (i.e. NA/DK answer) we provide five imputed values. These imputations are stored as five distinct datasets (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)].

To make inferences from the five multiply imputed datasets one has (1) first to analyze each of the five datasets 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 datasets we have obtained an estimate of Q (using standard complete-data methods), denoted $\hat{Q}_i$. The MI point estimate of Q, $\overline{Q}$, is the average of the five complete data estimates

$$\overline{Q} = \frac{1}{5}\sum_{i=1}^{5}\hat{Q}_i$$

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

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

$$W = \frac{1}{5}\sum_{i=1}^{5}\hat{V}_i$$

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

$$B = \frac{1}{4}\sum_{i=1}^{5}(\hat{Q}_i - \overline{Q})^2$$

The total variance for $\overline{Q}$ 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) if only means or similar statistics are of interest, an alternative to analyzing separately the five datasets and combining the results is to construct a dataset containing the five imputed datasets successively (i.e. a unique dataset where the number of observations is five times the actual number of respondents), divide the weight variable (*facine3*) by five, and calculate the statistic.

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

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

(iv) 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.

```
------------------------------------------------------------------------------
                    *OVERALL ESTIMATES (option (iv) above);
use c:\input.dta;
*the file input.dta should contain five observations and two variables which are the point estimate (called here bmean) and the standard error (called here bsemean) for each of the five datasets;
gen ni=5;
set type double;
gen varmean=bsemean*bsemean;
egen w=mean(varmean);
egen qbar=mean(bmean);
gen dev=(bmean-qbar)*(bmean-qbar);
egen be=sum(dev);
replace be=be*(1/(ni-1));
gen totvar=w+(1+(1/ni))*be;
gen sqrttotvar=sqrt(totvar);
* qbar denotes the overall point estimate, totvar the overall variance (within and between component), and sqrttotvar the overall standard error;
format qbar totvar sqrttotvar %12.1f;
list;
------------------------------------------------------------------------------
                    *USING MI STATA COMMANDS (option (iii) 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 effe_2017_imp1_dta.zip, value 2 if it comes from effe_2017_imp2_dta.zip 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 EFF 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 EFF data multiply imputed as follows: use eff; mi import flong, m(mdataset) id(h_2017), assuming *eff* is the name of this pooled dataset.[10]

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. GENERATE THE UNIQUE DATA SET;
*First dataset;
use databol1;
gen byte mdataset=1;
save eff, replace;

*Remaining datasets;
forvalues m=2/5 {;
  use databol`m', clear;
  gen byte mdataset=`m';
  append using eff;
  save eff, replace;
};

*Dataset with value 0 in the indicator of multiply imputed dataset;
use databol1;
gen byte mdataset=0;
append using eff;
save eff, replace;


*B. COMBINING RESULTS USING MI COMMAND;
mi import flong, m(mdataset) id(h_2017);
mi estimate, esampvaryok post: qreg renthog [pweight=facine3];
------------------------------------------------------------------------
```

6 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.

[10] 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.

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 households from within each cluster) is usual sampling practice in order to reduce costs but it may diminish precision if household 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 EFF data without disclosing stratum or cluster information we provide 1000 replicate weights[11]. 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 variance for the first implicate data set, i.e. V_1 in the notation of the previous section[12].

[11] These are the variables wt3r_*i*, wtpan1r_*i*, wtpan2r_*i*, *i*= 1, …, 1000 in the replicate weights files, as described in sub-section 1.2.
[12] This should be repeated for the rest of the implicates to obtain W .

```
-------------------------------------------------------------------------
* STANDARD ERRORS USING REPLICATE WEIGHTS (FOR THE FIRST IMPLICATE);
* HERE, FOR EXAMPLE, FOR THE MEDIAN;

* A.- Statistic of interest: original sample weighted median of the variable riquezanet;
* To calculate the statistic of interest we use here the Stata procedures described in Carlin et al. (2003);
miset using c:\databol;
mido pctile medvivpr=riquezanet [pweight=facine3];
mido list medvivpr in 1/2;
mido drop if _n>1;
mici, indiv: medvivpr;
clear;

* B.- OBTAINING THE STANDARD ERROR (FOR ONE OF THE FIVE IMPLICATE DATA SETS).
* FIRST IMPLICATE;
* We first merge our data with the replicate weights file;
use c:\databol1;
sort h_2017;
merge h_2017 using c:\replicate_weights_2017;
tab _merge;
drop _merge;
save c:\eff1wdata.dta;

* First bootstrap sample and its weighted median;
pctile medhp=riquezanet [pweight=wt3r_1];
list medhp in 1/2;
keep medhp;
drop if _n>1;
save c:\loop1, replace;
clear;

* Reps-1 bootstrap samples and their weighted medians;
set output error;
forvalues s=2/1000 {;
use c:\eff1wdata.dta;
pctile medhp=riquezanet [pweight=wt3r_`s'];
keep medhp;
drop if _n>1;
append using c:\loop1;
save c:\loop1, replace;
drop _all;
          };
set output proc;
use c:\loop1;
*The summarize command will provide the sampling standard error of the median in the first imputed data set, $(V_1)^{1/2}$;
```

```
sum;
clear;
-------------------------------------------------------------------------
```

For the most common estimation commands, Stata users can also make use of the svy command to estimate the variance, V_i, 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 eff1wdata.dta, and to specify the replicate weights using the svyset command.

Below we show an alternative way of estimating the variance, V_1, in the first dataset for the sample weighted mean of the variable analyzed in the example above.

```
-------------------------------------------------------------------------
* STANDARD ERRORS USING REPLICATE WEIGHTS (FIRST IMPLICATE): SVY COMMAND;
* HERE, FOR EXAMPLE, FOR THE MEAN;

* A.- Statistic of interest: original sample weighted mean of the variable riquezanet;
* To calculate the statistic of interest, we estimate a linear regression of the variable of interest on an intercept to recover the weighted sample mean from its estimated coefficient;
use eff;
mi import flong, m(mdataset) id(h_2017);
mi svyset [pweight=facine3];
mi estimate, esampvaryok post: svy: reg riquezanet;
clear;

* B.- OBTAINING THE STANDARD ERROR (FOR ONE OF THE FIVE IMPLICATE DATA SETS).
* FIRST IMPLICATE;
* We first merge our data with the replicate weights file as indicated in the example above for the median (generate the dataset eff1wdata);
use eff1wdata;
svyset [pweight=facine3], bsrweight(wt3r_*) vce(bootstrap);
svy: reg riquezanet;
*This operation should be repeated separately with the remaining implicate datasets in order to obtain W ;
-------------------------------------------------------------------------
```

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. GENERATE THE UNIQUE DATASET;
*Dataset with value 0 in the multiply imputed dataset indicator;
use eff1wdata;
gen byte mdataset=0;
save effwdata;

*Multiply imputed datasets;
forvalues m=1/5 {;
  clear;
  use eff`m'wdata;
  gen byte mdataset=`m';
  append using effwdata;
  save effwdata, replace;
};

*B. COMBINING RESULTS USING MI COMMAND;
mi import flong, m(mdataset) id(h_2017);
mi svyset [pweight=facine3], bsrweight(wt3r_*) vce(bootstrap);
mi estimate, esampvaryok post vceok: svy: reg riquezanet;
-------------------------------------------------------------------------
```

The dataset effwdata is the result of appending the five datasets effw1data,..., effw5data, 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.

REFERENCES

BANCO DE ESPAÑA (2019a). 'Survey of Household Finances (EFF) 2017: Methods, Results and Changes since 2014', Analytical Article*, 19 December 2019.*

BANCO DE ESPAÑA (2019b). 'Encuesta Financiera de las Familias (EFF) 2017: Métodos, Resultados y Cambios desde 2014', Artículo Analítico*, 19 diciembre 2019.*

BARCELÓ, C. (2006). Imputation of the 2005 Wave of the Spanish Survey of Household Finances (EFF), *Occasional Paper Nº 0603, Banco de España*.

BARCELÓ, C., L. CRESPO, S. GARCÍA-URIBE, C. GENTO, M. GÓMEZ, and A. DE QUINTO (2020). The Spanish Survey of Household Finances (EFF): Description and Methods of the 2017 Wave, *Occasional Paper Nº 2033, Banco de España.*

BOVER, O. (2004). The Spanish Survey of Household Finances (EFF): Description and Methods of the 2002 Wave*, Occasional Paper Nº 0409, 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.

APPENDIX:

VARIABLES THAT HAVE NOT BEEN IMPUTED

The variables for which no imputation is provided for NA/DK answers are the following:

1) p2_1e
2) p6_55b
3) p6_57b
4) p6_59bc1
5) p6_59a
6) p6_59b
7) p6_51a
8) p6_51b
9) p7_6b
10) p8_5c

For these variables, observations whose values should have been imputed but imputation was judged not reliable are marked with a -9999 value.