ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Household Finance and Consumption Network

Working Paper

The Household Finance and Consumption Survey: methodological report for the second wave

ECB Statistics Paper, No. 17

Provided in Cooperation with:
European Central Bank (ECB)

Suggested Citation: Household Finance and Consumption Network (2016) : The Household Finance and Consumption Survey: methodological report for the second wave, ECB Statistics Paper, No. 17, ISBN 978-92-899-2643-0, European Central Bank (ECB), Frankfurt a. M., https://doi.org/10.2866/260900

This Version is available at:
https://hdl.handle.net/10419/154652

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

EUROPEAN CENTRAL BANK

EUROSYSTEM

Statistics Paper Series

Household Finance and Consumption Network

The Household Finance and Consumption Survey: methodological report for the second wave

No 17 / December 2016

Disclaimer: This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

Household Finance and Consumption Network

This report was drafted by the Household Finance and Consumption Network (HFCN). The HFCN is chaired by Ioannis Ganoulis (ECB) and Oreste Tristani (ECB), and Sébastien Pérez-Duarte (ECB) and Jiri Slacalek (ECB) as Secretaries.
The HFCN is composed of members from the:
European Central Bank,
Banque Nationale de Belgique,
Česká národní banka,
Danmarks Nationalbank,
Deutsche Bundesbank, Eesti Pank,
Central Bank of Ireland,
Bank of Greece,
Banco de España,
Banque de France,
Hrvatska narodna banka,
Banca d'Italia,
Central Bank of Cyprus,
Latvijas Banka,
Lietuvos bankas,
Banque centrale du Luxembourg,
Magyar Nemzeti Bank,
Central Bank of Malta,
De Nederlandsche Bank,
Oesterreichische Nationalbank,
Narodowy Bank Polski,
Banco de Portugal,
Banca Naţională a României,
Banka Slovenije,
Národná banka Slovenska,
Suomen Pankki,
Sveriges Riksbank,

as well as
Statistics Estonia,
Central Statistics Office (Ireland),
Insee (France),
Hungarian Central Statistical Office,
Instituto Nacional de Estatística (Portugal),
Statistics Finland,
European Commission (Eurostat)

and consultants from the
Federal Reserve Board,
Goethe University Frankfurt
and University of Naples Federico II.

The HFCN collects household-level data on households' finances and consumption in the euro area through a harmonised survey. The HFCN aims at studying in depth the micro-level structural information on euro area households' assets and liabilities.
Email: hfcs@ecb.europa.eu

Contents

1 Introduction 3

1.1 General features of the HFCS 3

2 The HFCS blueprint questionnaire 10

2.1 Pre-interview part of the HFCS questionnaire 10

2.2 Topics covered by the HFCS core questionnaire 12

2.3 Interview closure and post-interview debriefing/paradata 16

2.4 Data collection approaches incorporated into the questionnaire 17

2.5 The HFCS non-core questions 18

3 Collection of data and other fieldwork aspects 20

3.1 Survey mode 20

3.2 Fieldwork 22

3.3 Deviations from the data collection framework: other data sources 24

4 Sample design 26

4.1 General features 26

4.2 Main country features 27

5 Unit non-response and weighting 37

5.1 Unit non-response in wealth surveys 37

5.2 Unit non-response in the HFCS 38

5.3 Weighting 39

6 Editing, item non-response and multiple imputation 44

6.1 Data editing 44

6.2 Imputation of the HFCS data 46

6.3 Comparative information on item non-response and imputation 52

7 Variance estimation 56

7.1 Motivation for replication-based methods 56

7.2 The Rao-Wu rescaled bootstrap and its extensions 57

7.3 Combining replicate weights and multiple imputation 61

7.4 Variance estimation of changes between waves 62

7.5 Software routines for estimating total variance 62

8 Statistical disclosure control 65

8.1 General principles in the HFCS 65

8.2 Collapsing of cases 67

8.3 Random rounding 67

9 Comparability issues 69

9.1 What are comparability issues? 69

9.2 Dimensions in the assessment of comparability 69

10 Comparability between the HFCS and other statistics 77

10.1 Comparability of the demographic structure of the HFCS 77

10.2 Comparing the HFCS and macro data on financial wealth and liabilities 82

10.3 Comparison of income data between the HFCS and EU-SILC 86

Appendices 90

HFCS definitions of financially knowledgeable person and HFCS household definition 90

Changes in the core variables for wave 2 92

Coverage of the core items in the second wave of the HFCS 94

Collection of the non-core items 97

Statistical disclosure: additional information 101

Revisions to data from the first wave 103

References 105

1 Introduction

1.1 General features of the HFCS

1.1.1 The Household Finance and Consumption Network

At the end of 2006, the ECB Governing Council set up the Household Finance and Consumption Network (HFCN). The network is composed of researchers, statisticians and survey specialists from the ECB, European national central banks (NCBs), some national statistical institutes (NSIs), and a number of experts in the field of household finances who act as consultants. The mandate given to the HFCN is to develop and conduct the Eurosystem Household Finance and Consumption Survey (HFCS), and act as a forum for research with the survey data.

While participation in the HFCN is purely voluntary, all euro area NCBs contribute to the HFCN and conduct the survey in their respective countries. In addition, several non-euro area NCBs participate as observers and, starting in the second wave, conducted the survey.[1]

1.1.2 General description of the HFCS

The HFCS is conducted in a decentralised manner. Each institution participating in the HFCN (NCB or NSI) is responsible for conducting the survey. The European Central Bank (ECB) in conjunction with the HFCN coordinates the whole project, ensuring the application of a common methodology, pooling and quality-controlling the country datasets, as well as disseminating the survey results and microdata through a single access gateway.

The HFCS is conducted every three years in most countries.[2] The fieldwork for the second wave was carried out in most countries between 2013 and the first half of 2015[3]. Table 1.1 provides a summary snapshot of the institution responsible for the HFCS in each country and the fieldwork periods.

---

1 The first wave of HFCS was conducted in 15 euro area countries and the second wave in 18 euro area countries, as well as in Hungary and Poland; the survey will be conducted in all euro area member states, including Lithuania, as of the third wave of the survey.

2 The HFCS is carried out every two years only in Italy.

3 Except in Spain, where the survey was conducted in 2011 and 2012.

Table 1.1

Main features of the HFCS country surveys

Country	Responsible institution	Fieldwork period	Adaptation of an existing survey*
Belgium	National Bank of Belgium	June-2014 - January-2015	HFCS
Germany	Deutsche Bundesbank	April 2014 - November 2014	HFCS
Estonia	Eesti Pank	March 2013 - June 2013	-
Ireland	CSO/Central Bank of Ireland	March 2013 - September 2013	-
Greece	Bank of Greece	June 2014 - October 2014	HFCS
Spain	Banco de España	October 2011 - April 2012	Yes
France	Insee/Banque de France	October 2014 - February 2015	Yes
Italy	Banca d'Italia	January 2015 - June 2015	Yes
Cyprus	Central Bank of Cyprus	February 2014 - July 2014	HFCS
Latvia	Latvijas Banka	April 2014 - September 2014	-
Luxembourg	Banque centrale du Luxembourg	April 2014 - December 2014	HFCS
Hungary	Magyar Nemzeti Bank	October 2014 - November 2014	-
Malta	Central Bank of Malta	January 2014 - June 2014	HFCS
Netherlands	De Nederlandsche Bank	April 2014 - March 2015	Yes
Austria	Oesterreichische Nationalbank	June 2014 - February 2015	HFCS
Poland	Narodowy Bank Polski	January 2014 - February 2014	-
Portugal	INE Portugal/Banco de Portugal	March 2013 - July 2013	HFCS
Slovenia	Banka Slovenije	September 2014 - December 2014	HFCS
Slovakia	Národná banka Slovenska	February 2014 - April 2014	HFCS
Finland	Statistics Finland/Suomen Pankki	January 2014 - May 2014	Yes

Source: ECB – HFCS metadata.
* *Yes* indicates that the HFCS first wave was an adaptation of an existing national wealth survey. *HFCS* means that the country participated in the first wave of the HFCS, which was a new national wealth survey. - indicates that the second wave of the HFCS is a new wealth survey.

1.1.3 Methodological features of the HFCS

The HFCS is designed around a common set of methodological principles, which ensures the comparability of results.

Ex ante comparability through an output-oriented approach

When compared with other international initiatives on household wealth surveys,[4] one of the most distinctive features of the HFCS is that the country wealth surveys that are part of the project follow an ex ante harmonised methodology. In particular, all HFCSs provide survey variables according to a set of common definitions and descriptive features according to an output-oriented approach. Of this set of survey variables with common standards and definitions, the HFCS has defined a set of core output variables, which all countries report to the ECB. In addition, there is a set of standardised non-core extensions that countries may voluntarily collect, and which

[4] Such as the Luxembourg Wealth Study.

therefore also provide comparable output, but only for those countries that collect the information.

Conversely, substantial cross-country differences within Europe imply that obtaining comparable information sometimes requires different questions in each country, as well as a considerable amount of country-level expertise. In turn, questions in country surveys may be somewhat adapted to the specific circumstances and financial products available to consumers in each country. Nonetheless, a common blueprint questionnaire is the starting point for country questionnaires. Country surveys can also collect country-specific (i.e. not necessarily comparable) variables. These are not included in the HFCS dataset, but only in national datasets available from the national central banks and statistical institutes.

In countries where there was no existing survey prior to the launch of the HFCS, full output harmonisation is achieved from the start. Conversely, in the countries where a survey was already in place (namely Finland, France, Italy, Netherlands and Spain), full convergence is to be achieved gradually. In Cyprus and Portugal, the existing wealth surveys were discontinued and replaced by the HFCS with the first wave.

Where there is a pre-existing survey, the gradual convergence process implies that for the first two waves, a few variables result from combination/adaptation of the original survey variables. The degree of convergence was improved in all such countries between the first and the second wave.

Sample design

Household samples have been designed in each country to ensure both euro area and country representative results.[5] This is particularly important taking into account the relatively large cross-country heterogeneity of financial markets, banking regulations, pension systems and fiscal policies in the euro area.

More than 84,000 households were surveyed in the second wave, with varying samples sizes across countries (see Chapter 4 for further details), up from 62,000 in the first wave; at the same time, five countries (Estonia, Ireland, Latvia, Hungary and Poland) are covered in the second wave for the first time.

All HFCS country surveys have a probabilistic sample design. This means that each household in the target population has an ex ante defined non-zero probability of being part of the sample. A more exhaustive description of the sample designs applied in each country is provided in Chapter 3.

[5] The target reference population for national surveys is all private households and their current members residing in the national territory at the time of data collection. Persons living in collective households and in institutions are generally excluded from the target population.

Oversampling the wealthy

Wealth surveys typically pursue two competing objectives: on the one hand, representing the behaviour of "typical" individual households and, on the other hand, representing a substantial fraction of total wealth. For the former target, it is optimal that the sample proportionally represents the population as a whole. For the second objective, the sample should adequately represent total wealth. Since wealth distribution is highly uneven, a given level of precision would either require a rather large (and costly) sample or, if efficiently designed, a sample which should include a disproportionally high number of wealthy households.[6]

Given the unequal distribution of household wealth and the fact that certain financial instruments are almost exclusively held (and in large quantities) by the wealthiest households, using data from a purely random selection of units would yield a statistically inefficient estimate of the wealth distribution.

In addition, response rates have a clear non-random component, in that wealthier households tend to be more difficult to contact and less likely to respond.

Against this background, 15 out of 20 countries participating in the HFCS oversample the wealthy via different methods. The methodologies applied, as well as the effectiveness of the oversampling, are further analysed in Chapter 4. All in all, oversampling wealthy households increases precision. Additionally, it reduces non-response bias for estimates for the top of the distribution, if one assumes that the coverage of household wealth will improve as a result of oversampling. It also improves efficiency in the estimation of variables positively correlated with wealth.

Panel component

A panel component is defined as households that are, by design, interviewed in at least two waves of the same survey. The existence of a panel component improves the measurement of changes between different waves, because part of the data refers to the same units in both waves. As has been previously outlined, the HFCS brings together country surveys which have been in place for years with newly created surveys set up specifically for the HFCS project. Some of the surveys in the first group already have a panel component, while some of the others have also initiated or plan to set up a panel as of the second HFCS wave. Countries that currently have a panel component are Belgium, Germany, Spain, Italy, Cyprus, Malta and the Netherlands. France and Slovakia plan to have a panel component in the next HFCS wave; Finland, using the rotational sample of EU Statistics on Income and Living Conditions (EU-SILC), will have a sample if the next wave is three years or less before the current one. Estonia is planning to have a panel component in the future.

[6] See for instance Kennickell (2007) and HFCN (2009). Further bibliography available under HFCN (2009) and Sanchez Munoz (2011).

Survey mode

Survey information in the HFCS is mostly collected through Computer-Assisted Personal Interviews (CAPI), i.e. face-to-face interviews administered by an interviewer using a computer to record the replies provided by respondents. Further details on the specifics of each country survey are provided in Chapter 3.

Data editing and imputation

After the fieldwork is concluded, the institutions responsible for the respective HFCS country surveys start a thorough process of detecting and correcting possible mistakes in the data. Such quality checks aim to correct various kinds of inconsistencies, such as mistyped or erroneous answers (e.g. amounts or frequencies). To this aim, intensive use is made of the comments and the paradata provided by interviewers at the conclusion of each interview.[7]

When there is no straightforward correction (for instance, if information was erroneously collected because of a problem in the routing of the questionnaire), the presumably erroneous variables are coded as missing, with a special flag indicating that the value was set to missing during editing, and should be imputed during the imputation phase.

Imputation is the process of assigning a value to a variable when it was not correctly collected or not collected at all. Imputation does not create information, and is no substitute for collecting the information in the first place. However standard econometric tools can only deal with complete datasets. Therefore, imputing missing values is almost always a pre-requisite for being able to use the data.

For the HFCS, a multiple stochastic imputation strategy has been chosen. The HFCS dataset provides five imputed values (replicates) for every missing value corresponding to a variable entering the composition of household wealth, consumption or income. A detailed description of the imputation procedure applied in the HFCS is given in Chapter 6.

1.1.4 Continuous survey evaluation, the need for future research and the variance-bias trade-off

Although some surveys that have become part of the HFCS have a long history and an accumulation of research on different methodological survey-related aspects, most of the surveys do not, and the HFCS as a whole is entirely new. Thus, a body of knowledge will need to be built in order to understand more deeply the effects of the different methodological options taken by countries, and other comparability and quality issues on the survey results.

[7] For further details, see Household Finance and Consumption Network (2008b).

In the case of complex surveys like the HFCS, all steps of data production might influence statistical inference, produced using the final data set. All decisions made with regard to the construction of the questions asked, definition of the target population, sampling design, coverage, non-response, protocols for survey execution, survey mode, editing, imputation, weighting design, tools for variance estimation and all other steps of survey production may have an important influence on the bias and variance of estimates based on final data.

The HFCS was guided by harmonised principles and methodologies with regard to all steps of data production; nevertheless, these methods were not fully converged due to the variety of differences in country-specific situations and institutions, as well as different priorities.

In the case of survey execution protocols, there are important known dimensions of differences, which are recorded in this methodological report. As regards statistical processing, the HFCS established high-level frameworks and in some instances made fairly detailed prescriptions. But inevitably, there is room for interpretation and judgement, and the resulting variation has the potential to affect true bias, true uncertainty of estimates and the degree of true bias or uncertainty that is actually measured. Often, there is a trade-off between measured bias and uncertainty in choices made in statistical processing. While it may be very difficult to describe in detail the true values of bias or precision, given the currently available information, it is possible to give an indication of trade-offs of bias and uncertainty. There are trade-offs in several aspects of statistical processing, including adjustments for unit non-response and weighting, imputation, variance estimations procedures, and other areas.

It should therefore be taken into consideration that datasets based on a data production process in which substantial variance was traded against bias will more often deliver “significant” results, even though they may have a larger true bias, which cannot be measured.

The HFCS is based on a strategy of transparency, allowing researchers to investigate to a reasonable degree how different choices in the data production process might have influenced the survey estimates directly or through a bias-variance trade-off. Additionally, the HFCN is committed to a continuous process of survey evaluation, focusing on the underlying measurement process and on achieving further harmonisation of the methodological approaches across countries.

In addition to this report, some national central banks or statistical institutes publish more detailed information on the methodologies applied at the national level. Table 1.2 lists national methodological reports of countries where these are available, or reports on the survey results that include some details on methodologies applied at the national level. Some of these reports are currently available in national languages. In several other countries not listed below, these kinds of reports for the second wave will be available later.

Table 1.2

National information on methodologies available for the second HFCS wave

Country	Methodological report / National report on survey results
Belgium	Du Caju, Philip: La répartition du patrimoine en Belgique: premiers résultats de la deuxième vague de la Household Finance and Consumption Survey (HFCS). Revue Economique, National Bank of Belgium, September 2016.
Germany	Knerr, Petra, Folkert Aust, Nina Chudziak, Reiner Gilberg and Martin Kleudken: Methodenbericht: Private Haushalte und ihre Finanzen (PHF), 2. Erhebungswelle 2014.
Estonia	Meriküll, Jaanika and Tairi Room: The assets, liabilities and wealth of Estonian households: Results of the Household Finance and Consumption Survey. Eesti Pank, Occasional Paper Series, 1/2016.
Ireland	Central Bank of Ireland: Household Finance and Consumption Survey 2013.January 2015.
Spain	Bover, Olympia, Enrique Coronado and Pilar Velilla: The Spanish Survey of Household Finances (EFF): Description and Methods of the 2011 Wave. Banco de España, Documentos Ocasionales No. 1407.
Italy	Banca d'Italia: Supplements to the Statistical Bulletin: Household Income and Wealth in 2012.
Luxembourg	Banque Centrale du Luxembourg: Quel est le niveau de culture financière au Luxembourg? BCL Bulletin 2016/01, pages 56-59. Banque Centrale du Luxembourg: L'enquête sur le comportement financier et consommation des ménages - Résultats de la deuxième enquête. Bulletin 2016/02, pages 41-51.
Austria	Albacete, N., P. Lindner and K. Wagner. 2016. Eurosystem Household Finance and Consumption Survey 2014. Methodological notes for Austria. Addendum to Monetary Policy & the Economy Q2/16.
Poland	Narodowy Bank Polski: Zasobność gospodarstw domowych w Polsce. Aneks metodologiczny.
Portugal	Costa, Sónia: Financial situation of the households in Portugal: an analysis based on the HFCS 2013. Banco de Portugal Economic studies, Vol II, No. 4, pp. 15-56.
Slovakia	Cupák, Andrej and Anna Strachotova: Výsledky Druhej Vlny. Národná Banka Slovenska, Príležitostná študia, 2/2015.
Finland	Statistics Finland: Households' Assets

2 The HFCS blueprint questionnaire

The HFCS blueprint questionnaire consists of three separate parts: introduction, questionnaire sections on the nine topics with household-level and person-level questions and interview closure. While the target euro area output is specified in terms of core variables and harmonised definitions, national questionnaires can to some extent be adapted to national specificities. The blueprint euro area questionnaire provides the wording of individual questions in English, and is used by national survey questionnaires as a benchmark.

2.1 Pre-interview part of the HFCS questionnaire

2.1.1 Interview introduction and selection of main respondent

The HFCS blueprint questionnaire provides a script for establishing contact with the sampled household as well as some introductory information (on the importance of participating in the survey, measures to ensure data confidentiality, how the survey data will be used, etc.) that all interviewers are instructed to read out to the interviewees before the start of the interview.

An important part of the interview introduction is the selection of the main household respondent, who is called the Financially knowledgeable person (FKP). The FKP is considered to be the main respondent, and provides financial information for the whole household, since this information is collected together for the whole household instead of by individual persons. This is to minimise response burden and to avoid duplications. For a survey like the HFCS whose main focus is on household finances, assets and liabilities, it is of vital importance to target the right person, so that the best available information on household finances can be collected during the interview.

The interview introduction contains a checklist of attributes providing detailed criteria on how to identify the FKP or, as a second best, the best available proxy, including provisions for special cases where the FKP is external to the interviewed household, for instance a relative outside the household (e.g. an independent child) taking care of the household's finances, a portfolio manager, an accountant, a lawyer or a tax adviser.[8]

[8] Further details on the selection script of the FKP are provided in the Appendix.

2.1.2 Household listing, HFCS household definition and reference person

The purpose of this part of the questionnaire is to establish a list of household members, i.e. defining the perimeter of the household. The replies of the main respondent regarding the household's financial information (assets, debts, consumption, etc.) should thus (only) refer to the household members identified in this initial step.

For the definition of household, the HFCS uses a variation of the so-called "housekeeping concept".[9] A household is defined as a person living alone or a group of people who live together in the same private dwelling and share expenditures, including the joint provision of the essentials of living.

Persons usually resident, but temporarily absent from the dwelling for a period of less than six months (for reasons of holiday travel, work, education or similar) are included as household members. Persons financially dependent and not having their private household somewhere else (like students studying away from home, persons away for work regularly returning and considering the sampled dwelling as their main place of residence) are included as household members even if their length of absence may exceed six months. Conversely, possible other persons with usual residence in the dwelling but who do not share expenditures (e.g. lodgers, tenants, etc.) are treated as separate households. Consequently, in some specific cases, there can be more than one household in a dwelling, but only a single household would be interviewed in that case.[10]

The outcome of the screening part is the list of household members verified against the household membership definition. Individual members are then listed according to their relationships with an interview reference person chosen from among the household members. The interview reference person may be, but need not always be, identical to the FKP. For instance, when the financial information for the household is provided by a person who does not belong in the household (an accountant, a lawyer, a grown-up child, etc.), the FKP and the interview reference person are necessarily different.

Additionally, the interview reference person defined at the beginning of the interview (i.e. the person around whom the household is drawn) may not be the same as the reference person used in the presentation of survey results. For instance, to release/tabulate survey results for some characteristics such as age, education or work status that can be assigned only at individual person level, one person must represent the household as a whole. Such a person must be chosen with pre-defined objective criteria, as the household will be classified according to this reference person's characteristics. The information necessary to apply a set of criteria is not yet available when the interviewer is asked to list the members of the household.

---

[9] As opposed to the dwelling concept, where all persons living in one dwelling are automatically considered as one household. See, for example, UN (2008), p.100 for a more indepth discussion of these two concepts.

[10] The complete household definition applied for the HFCS is provided in the Appendix.

The reference person for statistical outputs is therefore constructed ex post, based on all the information collected about the household during the interview.

In HFCS publications showing euro area results, the criteria are based on recent international standards for household income statistics presented by the so-called Canberra Group (UNECE, 2011). It uses the following sequential steps to determine a unique reference person in the household:

- one of the partners in a registered or de facto marriage, with dependent children,
- one of the partners in a registered or de facto marriage, without dependent children,
- a lone parent with dependent children,
- the person with the highest income,
- the eldest person.

2.2 Topics covered by the HFCS core questionnaire

The HFCS questionnaire is split into nine sections marked letters A to I, in addition to pre- and post-interview sections. The sections on demographics, employment, and pensions and life insurance policies cover information collected at the personal level, i.e. individually for all persons aged 16 or over. The sections on real assets and their financing, other liabilities and credit constraints, private businesses and financial assets, intergenerational transfers and gifts and consumption and saving cover questions/information collected at the household level. In the section on income, some income components are collected at the personal level (e.g. employment-related income, pension income, etc.) and some at the household level (e.g. income from financial investments).

Changes to the questionnaire between the first and second waves are listed in Box 2.1 at the end of Chapter 2.

2.2.1 Demographics

The demographics section contains a basic set of information collected for all household members, namely age, gender, country of birth, and length of stay in the country (for the foreign born). Information on marital status and the highest level of education attained are only collected for household members aged 16 or over.

2.2.2 Real assets and their financing

This section collects information on ownership and current values of real estate assets (household main residence for homeowners, other real estate properties owned by the household), vehicles (cars, other types of vehicles such as motorbikes, boats, etc.), valuables (such as jewellery, works of art, antiques) and a residual item for other real assets. A question is also asked on the purchase of vehicles within the past 12 months. Questions about other characteristics are asked for the household main residence (way and year of acquisition, value at the time of acquisition, etc.). Both owners and tenants are asked about the size of the household main residence and the length of stay in the current household main residence. Tenants also provide information about the monthly amount paid as rent. For other real estate properties, the type of owned property, its main use (for private use/for own business/for rent), the percentage of the property owned by the household and its current value are asked in a loop for up to three main properties.

A collection approach that asks for mortgages by collateral is applied in the HFCS questionnaire. After the questions on the household main residence, a set of questions is asked on the characteristics of each mortgage collateralised by the property. The same approach is followed with other real estate properties, i.e. questions referring to each mortgage collateralised by other real estate properties are asked immediately after information is collected about the properties. This reduces the risk of respondents forgetting to report on specific debts.[11]

Selected details containing purpose of the loan, year when the loan was taken out or last refinanced, initial amount borrowed, initial maturity, current interest rate, whether the interest rate is fixed or adjustable, and current monthly payment made on the loan are asked in loops for up to two or three mortgages on the household main residence and up to three mortgages on other real estate properties.

2.2.3 Other liabilities, credit constraints

The section on other liabilities contains questions on non-mortgage debt instruments – leasing contracts, credit lines/overdrafts, credit cards, private loans from family or friends and other loans not collateralised by real estate. On loans not collateralised by real estate, a loop for up to three main loans collects individual details such as the purpose of the loan, initial amount borrowed, initial maturity, current outstanding amount, current interest rate and current monthly payments. The remaining part of the section targets questions on loan application (applied for credit in the last three years) and credit constraints (credit refusal experience, not applying for credit due to perceived credit constraint).

[11] Some of the HFCS countries (Italy, Spain, France and Finland) use a different data collection approach in their national questionnaires, asking loans by their main purpose and then assigning them to collaterals. Data in these countries are output harmonised and recoded into the HFCS variables scheme using the per-collateral approach.

2.2.4 Private businesses, financial assets

The first part of this section covers self-employment private businesses (with the loop for details on up to the three most important: sector of activity [NACE[12]], legal form, number of employees, household members working in the business, share of the business owned by the household and the current value of the household's share in the business). These are distinguished from other "passive" investments in non-publicly traded equity, for which only questions on ownership and on total current value of the equity holdings are asked.

The second part then covers financial assets: sight accounts, saving accounts, mutual funds, bonds, publicly traded shares, additional assets in managed accounts, money owed to the household, and a residual question on other financial assets. Selected additional questions are asked for bonds (type of bonds owned – government/banks and financial corporations/non-financial corporations), mutual funds (type of mutual funds owned and current value of investments by type of mutual fund) and shares (ownership of foreign shares). The section also includes a self-assessment question on risk attitudes.

2.2.5 Employment

Employment section questions are asked to all household members aged 16 or over. The first question asks for the self-reported current labour status of each person. Persons in employment are asked a set of questions on the main characteristics of their employment: employment status (employee/self-employed/unpaid family worker), occupation (ISCO[13]), sector of activity (NACE), permanent/temporary contract for employees, hours worked per week, length of employment in the firm/with current employer, question on secondary employment activities in addition to the main job, expected retirement age. Those currently not in employment are asked a question on previous full- or part-time work. All employed persons or those with previous employment activity are asked about the total length of their employment. All persons who are not yet retired and are currently or have in the past been employed are asked a question on the age they plan to stop working.

2.2.6 Pensions and life insurance policies

The HFCS classifies pension wealth as voluntary pension schemes and life insurance contracts, occupational pension plans and public pension plans. Voluntary pension schemes and life insurance contracts are included in households' financial wealth in the report of HFCS results. The part on public and occupational pension plans aims to collect basic information on participation of household members aged 16 or over in these types of pension plans, and on the current value of plans with an

[12] See details of the NACE classification.

[13] See http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm for details of the ISCO classification.

account balance, if known to the respondent. This particular part of the questionnaire is labelled as indicative, open to particular national implementations. For instance, in the Netherlands and Finland, so-called defined benefit schemes for occupational pensions are significant components of household wealth, but do not correspond with the definition of occupational pension schemes with an account balance. Therefore, a (non-core) variable on the current value of all occupational plans that do not have an account has been produced in these two countries.[14]

2.2.7 Income

The HFCS is a survey focused on the collection of information on household wealth. Therefore, the main target of the income section is the collection of the main components for the construction of total gross household income, not including lower level details of each of these components (such as, for example, the further breakdown of income from financial assets).

This section combines personal-level questions (employee income, self-employment income, income from public pensions, income from private and occupational pensions, unemployment benefits) and household-level questions (social benefits other than pensions and unemployment benefits, regular private transfers received, rental income, financial investments income, private business or partnership income, other residual sources of income).

The concepts and definitions of the income section were designed along the lines of those of the UNECE Canberra group handbook on household income statistics.[15] Imputed rents and income in kind components are not covered by the HFCS core income section. The target income aggregate is gross, including taxes and social insurance contributions paid by employees.[16]

The reference period is 12 months, which could either be the last calendar year or the 12-month period preceding the interview, depending on the circumstances in individual countries. The last calendar year was used as a reference period in 16 countries; Ireland, Greece, Cyprus and Hungary used the past 12 months as a reference period for income.

In addition to the income-component questions, two qualitative supplementary questions are asked on the level of annual income as compared with normal and on income expectations over the following year.

[14] This variable is included in wave 2 to enable the use to adjust for the otherwise distorted net median and mean wealth position of Dutch and Finnish households in comparison with other countries.

[15] UNECE (2011).

[16] There are some cross-country differences in the strategies to collect information on income (see Chapter 9.2.5 for details).

2.2.8 Intergenerational transfers, gifts

This section collects information on received inheritances and substantial gifts, and is aimed at tracing household wealth accumulation patterns. The loop for up to the three most important transfers and gifts contains questions on when they were received, what asset types were received, their value and from whom they were received. The section also includes a question about expected substantial gifts and/or inheritances.

2.2.9 Consumption and saving

This section focuses on selected aspects of household consumption and saving. It collects information on several consumption indicators that, according to the literature,[17] may be used to infer total consumption. These items are spending on food at home, spending on food outside the home and spending on utilities. Additionally, one item on overall spending on all consumer goods and services is collected. All consumption items refer to spending in a typical month.

In addition, collected items include regular private transfers made outside the household (alimony, assistance, etc.), saving motives, comparison of last 12 months' expenditure with the usual level (higher/normal/lower), balance of expenditures and income (expenses higher than/equal to/lower than income) and ability to get emergency (financial) assistance from friends or relatives.

2.3 Interview closure and post-interview debriefing/paradata

The last part of the questionnaire covers one question intended to close the interview on topics and items that the respondent may have forgotten to report before.

After the interview, an additional set of questions is aimed at collecting feedback from interviewers (so-called paradata). The interview paradata section encompasses 16 questions covering aspects surrounding the interview, e.g. the accuracy of the respondent's calculations, who was present during the interview, perceived trust of the respondent before and after the interview, etc. This information is deemed very valuable for the treatment of the data ex post, i.e. for data editing and imputation.

Box 2.1
Changes from the wave 1 questionnaire

The introduction to the questionnaire, in which the FKP is identified and all members of the households are listed, was simplified after the first wave to allow more flexibility in how to attain the required information. The structured set of questions was replaced by a shorter checklist of

[17] See for example Browning, Crossley and Weber (2003).

attributes. It was also highlighted that countries are not meant to strictly implement the instructions, but the focus should be on the comparability of the output.

The questions block on loans was reorganised to make a clearer distinction between loan refinancing (i.e. paying off an existing loan with the proceeds from a new loan, allowing the borrower to benefit from better terms) and loan renegotiation (i.e. a change in some of the original loan parameters, either upon request from the borrower or stemming from periodic revision of selected loan parameters that involve negotiation).

A new set of questions on informal loans from family or friends (private loans) was added to make a clearer distinction between these kinds of loans and other non-collateralised loans. In the first wave, such loans were collected as a part of the item “other non-collateralised loans”. A question was asked about the number of private loans, as well as the purpose and outstanding balance of the three private loans with the largest outstanding balances.

New questions were added on the purchase of cars or other vehicles in the past 12 months.

The question on whether any lender or creditor has turned down any request was modified to include the possibility of multiple answers. The household can now report both having their credit application turned down and not being given as much credit as they applied for.

Two additional questions on household consumption expenditure were added. The first question was on the consumption of utilities (electricity, water, gas, telephone, internet and television), and the second on the consumption expenditure on all consumer goods and services.

In addition, based on the experiences from the first-wave data collection, some clarifications in question wordings and interviewer instructions were made to improve the quality of collected data. These additions did not have any impact on the definitions of output variables. A few changes were introduced to the filtering of individual questions, to avoid unnecessary data collection (e.g. question "At what age do you plan to stop working for pay?" will not be asked if the respondent reported never having been employed).

Two questions were dropped from the interview closure. The first was on questions that the respondent found especially difficult to answer, and the second on any suggestions or comments to the interview.

2.4 Data collection approaches incorporated into the questionnaire

2.4.1 Loops

Loops are sequences of questions referring to individual items, which are repeated for each individual item. There are seven loops in the HFCS core questionnaire, collecting details on household main residence mortgages, other real estate properties, mortgages on other real estate properties, private loans, non-

collateralised loans, self-employment businesses and gifts/inheritances received. Each loop sequence starts with a question on the number of instances (e.g. number of loans, number of other properties) followed by a set of questions on details which are repeated for up to three main items. The loop ends with a mop-up question collecting aggregate information on remaining items four and above, for which details are no longer collected (e.g. the total outstanding amount for loans number four and higher, properties).[18]

2.4.2 Collection of monetary value questions

A standardised CAPI data collection script is used to collect monetary values (called the "Euroloop", as it targets the collection of values in euro, or in national currencies in non-euro area countries). The Euroloop encompasses a set of questions which should be asked in a strict sequence.

First, the interviewer should ask the exact amount, which respondents may provide either in euro or in national legacy currencies. Only if respondents are unable (or unwilling) to provide the exact amount should the interviewer then proceed to ask the respondent to provide the information in flexible brackets, i.e. to provide self-reported upper and lower bounds. If the respondent is still unable to answer, there is a third step involving a card with 20 prefilled fixed intervals in euro and corresponding amounts in national legacy currencies. In this last step, the coded amount or interval (lower-upper bound) are displayed to the respondent as numbers and spelled out to check and confirm.

After collecting each reply, interviewers are instructed to repeat aloud the amount reported by respondents in order to try to correct possible mistakes on the spot.

2.5 The HFCS non-core questions

The blueprint questionnaire covers the core HFCS variables. In addition to the core survey content, the HFCN prepared a supplementary harmonised set of non-core variables, which usually supplement the topic covered by the existing core questionnaire parts with more detailed information. The HFCS non-core part also includes one additional section on payment habits.

The recommended question wording and the recommended position in the questionnaire vis-à-vis the related core survey items are provided in the HFCS non-core variables catalogue. This provides a guideline as to how the non-core questions can be inserted into the core national questionnaires.

[18] In some countries, simplified loops of up to two items with a mop-up question for items three and above are used.

By their nature, non-core variables are collected only in a subset of the HFCS countries. An overview of non-core variables covered in one or more of the HFCS country files in wave 2 is provided in the Appendix.

3 Collection of data and other fieldwork aspects

The HFCS data collection is ex ante output harmonised with a list of core output variables that every country should collect in accordance with a set of common definitions. However, the HFCS output harmonisation enables a few temporary deviations from the recommended data collection mode and the use of other reliable data sources complementing/completing the survey data, over a transitory convergence process encompassing one or several survey waves. In addition to data collection, various other fieldwork issues are also examined in this chapter.

3.1 Survey mode

The type of interaction between the respondent and the survey questionnaire is an important determinant of possible measurement error. The first and most important decision for a household survey is therefore the selection of the mode of data collection (Jäckle, Roberts and Lynn, 2006; Dillman and Christian, 2005). Using different modes to interview different sample units entails a high risk of comparability between survey results (de Leeuw, 2005). In a multi-national setting, this risk also becomes evident in comparisons between different countries using different survey modes.

For the HFCS, the same survey mode should be applied throughout all sample units in a country and across countries. The survey mode chosen for the HFCS is Computer Assisted Personal Interviews (CAPI), i.e. face-to-face interviews administered by an interviewer using a computer to record the replies provided by respondents. Survey data can be complemented by administrative data for variables with available consistent register sources. The use of a computer allows a smooth and error-free administration of the routing of the questions (which is particularly complex in the HFCS questionnaire), the application of consistency checks during the interview and the automatic storage of the data. Eliminating errors at the interview stage improves the quality of the survey data, and may save considerable resources in the subsequent data editing and cleaning phase.

In addition, interviewers play an important role in collecting high-quality income and wealth information, namely in: (1) persuading respondents to participate in the survey, increasing response rates, and reducing the risk of response bias; (2) building up trust vis-à-vis respondents, thus lowering the likelihood that a respondent will drop out in the middle of an interview; (3) minimising levels of item non-response by personally assisting (i.e. offering pre-designed prompts) – if required – during the interview; (4) avoiding incomplete responses; (5) providing additional information (interviewers' observations and paradata); etc. (HFCN, 2008a).

To a large extent, HFCS uses a single-mode approach within countries, meaning that there is one dominant survey mode in each participating country (see Table 3.1). For mainly practical reasons, a small share of interviews was conducted via a mode other than the dominant one in various countries, but this share is in most cases negligible. While 17 countries applied CAPI interviews in the second wave, in three countries, CAPI was not the main data collection method. In Poland, Paper-and-Pencil Interview (PAPI), in Finland, Computer Assisted Telephone Interview (CATI) and in the Netherlands, Computer Assisted Web Interview (CAWI) were the dominant survey modes. In Finland, most items on wealth, liabilities and income were not collected by interviews at all, but drawn directly or estimated with information from administrative registers. All countries that participated in the first HFCS wave used the same main survey mode in the second wave, except for Cyprus, where most of the interviews were conducted using PAPI in the first wave.

The median duration of the interview was in most countries slightly less than one hour. In most countries that conducted the first HFCS wave, the median interview duration was slightly longer in the second wave. This is to a large extent caused by the increase in the number of variables in the HFCS core variables list. The interview lengths are not directly comparable, since the numbers of questions and variables collected in different countries varied to some extent. Especially in countries in which the HFCS was a continuation of an existing wealth survey, a great deal of information from outside the core variable list of the HFCS was collected to maintain the time series of the national wealth surveys.

Table 3.1

Share of interviews by survey mode in HFCS countries and length of interviews

Country	CAPI	CATI	CAWI	PAPI	Median length of interview (minutes)
Belgium	100	0	0	0	58
Germany	100	0	0	0	66
Estonia	100	0	0	0	48
Ireland	100	0	0	0	41
Greece	100	0	0	0	46
Spain	100	0	0	0	60
France	100	0	0	0	75
Italy	92.9	0	0	7.1	47
Cyprus	100	0	0	0	60
Latvia	100	0	0	0	46
Luxembourg	100	0	0	0	56
Hungary	68.6	0	31.4	0	30
Malta	83	0	0	17	44*
Netherlands	0	0	100	0	-
Austria	100	0	0	0	70
Poland	0	0	0	100	70
Portugal	100	0	0	0	50
Slovenia	100	0	0	0	31
Slovakia	100	0	0	0	60
Finland	2.5	97.5	0	0	27†

Notes: CAPI: Computer Assisted Personal Interviews; CATI: Computer Assisted Telephone Interviews; CAWI: Computer Assisted Web Interview; PAPI: Paper-and-Pencil Interview.
* Excludes the screener, household listing and interview closure, as well as interviews conducted by PAPI.
† Refers to the Income and living conditions survey that included a module on household wealth and liabilities.

3.2 Fieldwork

In nine countries, the national statistical institute (NSI) was in charge of data collection, and interviews were conducted by staff in the survey units of the corresponding NSIs (see Table 3.2). In all other countries, the organisation responsible for conducting interviews was an external survey agency selected by the National Central Bank (NCB) in charge of the survey. In the Netherlands, a research institute was responsible for collecting the HFCS data through a web survey.

Interviewers were either employees of the survey agency or the NSI in charge of the data collection, or freelancers directly recruited by the survey agency. Before the start of the fieldwork, all countries organised face-to-face training sessions for interviewers. The training included both generic topics on how to motivate respondents to participate in a survey, as well as HFCS-specific issues explaining the concepts and definitions used in the wealth survey. Most training sessions also included practical exercises in which the interviewers had to conduct a test interview. All NCBs or NSIs in charge of data collection participated in the training sessions.

Fieldwork periods in the second wave of the HFCS varied from less than two months in Hungary and Poland to over ten months in the Netherlands. Shorter fieldwork periods are beneficial for data comparability, either because the reference periods for income or balance sheet items are closer or, in the case of a fixed reference period, to minimise recall bias. Conversely, longer fieldwork periods allow more opportunities to increase the number of contact attempts and thus obtain a higher number of interviews. The number of interviewers varied across countries, to a large extent depending on the sample size. The number of language versions of the questionnaire varied from one to five.

Table 3.2
Fieldwork indicators

Country	Organisation responsible for fieldwork	Number of interviewers conducting the survey	Language versions of the questionnaire	Length of fieldwork period (months)	Adaptation of existing survey (other than HFCS wave 1)
Belgium	SA	123	French, Dutch, English, German	7.5	N
Germany	SA	311	German, Russian, Polish, Turkish, English	7.5	N
Estonia	NSI	71	Estonian, Russian	4	N
Ireland	NSI	40	English	6	N
Greece	SA	69	Greek	5	N
Spain	SA	84	Spanish	7	Y
France	NSI	500	French	5	Y
Italy	SA	188	Italian, English	6	Y
Cyprus	SA	25	Greek, English	5	N
Latvia	NSI	48	Latvian, English, Russian	5.5	N
Luxembourg	SA	61	English, French, German	9	N
Hungary	NSI	262	Hungarian	1.5	N
Malta	SA	27	English, Maltese	5	N
Netherlands	SA	Not applicable	Dutch	10.5	Y
Austria	SA	72	German	9	N
Poland	NSI	695	Polish	1.5	N
Portugal	NSI	131	Portuguese	3.5	N
Slovenia	SA	32	Slovenian	4	N
Slovakia	NSI	128	Slovak, English	3	N
Finland*	NSI	140	Finnish, Swedish	4	Y

SA = Survey Agency, NSI = national statistical institute
* Parts of the data were collected from the EU-SILC survey, selection of target variables based on the HFCS and previous wealth surveys by Statistics Finland.

Of the 20 countries participating in the second wave of the HFCS, 15 had already conducted the first wave of the survey. In ten of these countries, the first wave of the HFCS was a new wealth survey, in most cases the first household wealth survey of any kind organised by the NCB. Three central banks added harmonised HFCS output variables to an existing wealth survey. These countries and their surveys were Italy (*Indagine sui Bilanci delle Famiglie Italiane* – Survey on Household Income and Wealth, SHIW), the Netherlands (DNB Household Survey, DHS) and Spain (*Encuesta Financiera de las Familias*, EFF). In France, the HFCS was a joint effort between the NCB and the NSI (Insee), and an adaptation of the *Enquête Patrimoine* previously conducted by Insee. In Finland, the survey was based on the variables

included in the former Statistics Finland's household wealth survey (*Kotitalouksien Varallisuustutkimus*), complemented by the additional variables included in the HFCS core output variables. In Portugal, the HFCS replaced the Household Wealth Survey (*Inquérito ao Património e Endividamento das Famílias*, IPEF), which was already a joint project of Banco de Portugal and Statistics Portugal (INE).

In the five countries (Estonia, Ireland, Latvia, Hungary and Poland) that did not participate in the first wave of the HFCS, the second wave of the survey was the first wealth survey organised by the NCB. In all five of these countries, the fieldwork was conducted by the NSI.

3.3 Deviations from the data collection framework: other data sources

The ex ante output harmonisation of HFCS data enables the use of data collection methods other than a survey, whenever they are considered to provide better quality. In most countries, though, most variables were collected through surveys. The main exception is the Finnish data, which draw on sample material from Statistics Finland's income and living conditions survey as well as numerous types of register data and estimation methods. In other countries, different data collection methods were used in the production of only a few individual variables. Additionally, for some variables, the production of the survey variables included various kinds of estimation methods. Collection of gross income is probably the most significant, with a variety of country differences in data collection, and is covered in Chapter 9 on comparability issues.

In several countries, information other than survey data was used to construct HFCS variables. In addition to Finland, several variables on financial wealth and income are taken from registers and data of financial intermediaries in Estonia. Income variables in France are based on tax files. Legislative information was used to construct some pension variables; these questions were left out of the questionnaire. A summary of the cases is shown in Table 3.3. Also, cases where register data were used are listed below for a complete coherence analysis. Register data are used in various other surveys to replace survey data, if the sources are reliable and the definitions of the register sources identical to the definitions of the corresponding target variables.

The variety of the estimation methods used by Statistics Finland to collect data on some wealth items was quite large. For example, the values of the main residence and other properties were formed by using data describing buildings and dwellings in the Population Information System and the data in the Tax Administration's housing company stock register. The values of vehicles were estimated based on data in several vehicle registers, price register systems and websites advertising boats for sale. Several variables on liabilities were constructed by combining information on tax registers and interview data. The values of unlisted shares were formed on the basis of dividend data obtained from individual taxation material. Of financial assets, pension wealth was estimated based on the individual tax register using the so-called perpetual inventory method. Deposits were only collected from households in

the third and fourth rotation group of the EU-SILC survey, and values for the first two rotation groups were constructed using statistical matching.

Table 3.3

Other data sources

Country	Information
Belgium, Germany, Greece, Netherlands	Legislative and institutional information is used to construct the percentage of current gross earnings contributed to the main public pension plan.
Estonia	Registers: income from public transfers, unemployment benefits and Estonian public pensions. Financial intermediaries: financial assets held in Estonia, II pillar pension funds. Combination of registers and interview data: mortgages, consumer loans from commercial banks, income from financial assets.
Ireland	Register data on income, including employee, profits and social transfers such as unemployment benefits and pensions etc., was used in the derivation of income.
France	Income data derived from administrative sources. Legislative information was used to construct some pension variables
Italy	Income from financial investments not directly collected, but calculated using average interest rates and information collected on households financial assets. Legislative and institutional information is used to define individuals' eligibility to receive public pensions.
Luxembourg, Slovenia, Slovakia	Information on the number of public pension schemes and the percentage of current gross earnings contributed to the main public pension plan are completed from the legislative and institutional parameters.
Latvia	Register data on real estate properties, credits and income were used to identify missing answers and to edit values of corresponding variables.
Finland	Register data: all income variables except private transfers and interest received, ownership and number of cars and other vehicles, business wealth, ownership and values of mutual funds, bonds and listed shares, and education. Estimated data: value of household main residence, ownership and value of other properties, values of cars and other vehicles, ownership and values of deposits, and values and contributions to voluntary pension schemes. Combination of registers and interview data: number, type and outstanding amounts of liabilities, loan payments.

4 Sample design

The comparison of sample designs is an essential part of evaluating how accurately the results of a survey represent the reality of its target population. This chapter analyses the main features of the sample designs and sampling frames chosen by the countries participating in the HFCS.

A vital point for wealth surveys is the efficiency with which information from the wealthiest part of the population is collected. This chapter provides a description of the approaches applied in different countries to oversample wealthy households.

4.1 General features

Sample design provides the most fundamental measurable statistical basis to evaluate a household survey. A good design should provide the most efficient and unbiased representation of the relevant population (Kennickell, 2005). Sampling design and implementation is a central component in the potential errors in estimation related to survey data (Verma and Betti, 2008), including errors on coverage, sample selection and also sampling errors and estimation bias.

The first and probably most important feature of the HFCS sample design is the use of probability sampling. This means that each household in the target population has a non-zero probability of being selected in the sample, and this probability should be known beforehand (HFCN, 2008a). Given the sizeable fixed costs of conducting a survey like the HFCS compared with the marginal costs corresponding to each additional sampling unit, the sample size should be representative both at the country and at the euro area level.

Since wealth is distributed very unequally, all participating countries are encouraged to explore methods for oversampling the wealthiest households.

Another relevant feature of the sample design for any survey is whether it is intended to introduce a panel component, i.e. whether (at least a portion of) the same households will be interviewed again over subsequent waves. In such a case, survey compilers need to take care to ensure the representativeness of both the cross-sections and the longitudinal component, and to ensure proper refreshment covering for sample attrition. All this may substantially add to the complexity of the sample design.

4.2 Main country features

While probability sampling is applied in all HFCSs in the second wave,[19] countries have adopted a variety of approaches in their sampling designs. The methodologies are largely dependent on the external data (population registers, postal addresses, dwelling registers, etc.) available for building up the sample.

4.2.1 Sampling designs applied

In household surveys, stratification of the population prior to sample selection is a commonly-used technique. In a stratified sample, various strata are constructed on the basis of auxiliary information that is known about the population, and sample units are selected independently from each stratum in a manner consistent with the survey's measurement objectives (UN, 2005). Units to be interviewed can be selected in one or multiple stages. In a multiple stage design, the first stage (or stages) involves a selection of geographical areas, from which individual households are chosen in the final stage.

Table 4.1 describes the sampling designs used in various countries. Five countries used one-stage stratified sampling, while 13 countries had a multi-stage stratified sampling design. In Malta and the Netherlands, no stratification was applied. In all countries, the sample size was chosen to be representative also at the country level.

Table 4.1
Sampling designs in the HFCS

Type of sampling design	Countries adopting
1-stage stratified sampling	BE, EE, CY, LU, FI
2-stage stratified sampling	IE, GR, ES*, FR, IT*, LV, HU, AT, PL, PT, SI, SK
3-stage stratified sampling	DE#
1- stage sampling	MT, NL

* In Spain and Italy, one stage for households living in municipalities with over 100,000 and 40,000 inhabitants respectively, two stages for others.
# In Germany, three stages for households living in municipalities with over 100,000 inhabitants, two stages for others.

Table 4.2 describes the stratification criteria in various countries. The sampling frames involved data on regions in the first stage (in multi-stage designs) and information on persons, households or dwellings in the second stage (or in the first stage in one-stage designs).

Region and population size of regional units were the most frequently used stratification variables, regions being in several cases additionally divided by the degree of urbanisation. Other stratification criteria included personal or regional average income, labour status and personal taxable wealth.

[19] In the first wave, probability sampling was used in 14 out of 15 countries; only Slovakia used quota sampling.

Table 4.2

Sampling frames and stratification criteria

Country	Sampling frame(s)	Stratification criteria
Belgium	National population register	Region, average taxable income by statistical sector and average dwelling price by municipality
Germany	List of street sections (Geodata); register of local residents from municipalities	Municipality size, anticipated wealth
Estonia	Population and housing census	Five NUTS3 regions and two income groups, the highest decile and the rest
Ireland	Population and housing census	Eight NUTS3 regions and five quintiles of deprivation/affluence.
Greece	List of municipalities, cities, villages and building blocks from census; dwellings	NUTS II region, degree of urbanisation
Spain	Population register supplemented with tax record information	Taxable wealth, municipality size
France	Tax register on main residences	Geographical area and common property
Italy	List of municipalities, population register	Municipalities by region and demographic size
Cyprus	Customer register of the electricity authority	Counties divided into urban and rural areas
Latvia	Population register, tax register; list of addresses	Degree of urbanisation (three groups), and income (three groups)
Luxembourg	Social security register	Nationality, employment status, monthly income
Hungary	Register of localities; register of dwellings	Regions, number of housing units, personal income tax
Malta	Population and dwellings register	Not applicable
Netherlands	CentERpanel	Not applicable
Austria	List of enumeration districts; register of post box addresses	Region (NUTS 3) and community size classes
Poland	Local data bank; Population and housing census	Regions (NUTS2), size of region, wealth (tax income and size of properties)
Portugal	National dwellings register	Nine regions (Nuts2 with disaggregation of Norte into Porto and Other Norte, and NUTS III for Lisbon).
Slovenia	Register of spatial units; Central population register	Demographic size of municipality
Slovakia	Household units database from census; database of occupied housing units	Eight regions (NUTS3) by three income groups
Finland*	Population information system of NSI	Income, type of income (personal taxable income of the main income earner of the household-dwelling unit).

Table 4.3 shows the numbers of strata used in the sampling designs of various countries. It also indicates the number of units, such as geographical areas or clusters, selected in the first stage in multi-stage designs (primary sampling units, PSU).

Table 4.3

Numbers of strata and primary sampling units selected

Country	Number of strata	Primary sampling units selected, for multi-stage designs
Belgium	24	-
Germany	4	200*
Estonia	10	-
Ireland	40	400
Greece	13	308
Spain	32	4116+985
France	22	567
Italy	53	371
Cyprus	8	-
Latvia	9	480
Luxembourg	20	-
Hungary	111	13596+84
Malta	-	-
Netherlands	703†	-
Austria	185	619
Poland	106	1786
Portugal	9	677
Slovenia	6	359
Slovakia	24	840
Finland	52	-

*Refers to the refresher component of the sample only.
†Refers to the first survey that was conducted to the panel.
Note: number of strata refers to the first sampling stage only. Primary sampling units selected are shown for countries with multi-stage sampling designs.

4.2.2 Panel component

In four countries (Spain, Italy, Netherlands and Finland) that adapted the HFCS to existing wealth surveys, a panel component was already in use. In Finland, the households participate in the annual income and living conditions survey for four consecutive waves. Consequently, the second-wave HFCS data, collected four years after the first wave, do not include any panel households. The French wealth survey introduced a panel dimension to its wealth survey in the second HFCS wave. France, Slovakia and Finland will have a panel component in the third HFCS wave. Four countries (Belgium, Germany, Cyprus and Malta) that conducted their first wealth survey in the first HFCS wave have a panel component in the second wave.

Table 4.4
Countries with a panel component

Country	Number of households re-contacted at wave 2, % of all contacted households	Panel design
Belgium	30	Pure panel with refresher sample
Germany	18	Pure panel with refresher sample
Spain	39	2002 all units included, 300 (600) units from 2005 (2008) randomly dropped
Italy	26	Units that participated in the previous wave are only randomly selected, all units that participated in the previous wave and at least one earlier wave are included
Cyprus	72	Pure panel with refresher sample
Malta	49	Pure panel with refresher sample
Netherlands	41	Pure panel with refresher sample

Source: ECB – HFCS metadata

In most of these countries, all units that were interviewed in the first wave were included in the gross sample of the second wave. In Italy and Spain, selected units from the first wave were excluded from the second-wave sample. The Spanish sample included households that had participated in the survey for two or more waves. In Italy, all households that had participated in the previous wave and at least one earlier wave were included in the second-wave sample. In Germany, all households that had participated in the first wave and agreed during the interview to have their address stored by the survey company for future waves (90% of the first-wave net sample) were re-contacted. In addition, refresher samples were introduced in all panel surveys to compensate for attrition. In most countries, the share of new households added to the sample was bigger than the share of panel households (see Table 4.4, third column).

4.2.3 Non-coverage of specific sub-populations in the sampling frame

The sampling frames of the HFCS included only households living in the countries where the survey was conducted. In addition, in most national surveys, the whole of the institutionalised population was left out of the sampling frame. Some other relatively small groups of the population are excluded from the sampling frames of individual countries. The gross sample of Cyprus did not include the population in Northern Cyprus.

Individuals belonging to some of the excluded groups, however, can be included in the sample, if they are considered as part of a household that is part of the sampling frame.

Table 4.5
Excluded groups

Country	Excluded groups
Belgium	Population in institutions (residents in homes for the elderly were included in the sampling frame), homeless
Germany	Population in institutions, homeless
Estonia	Population in institutions, persons with erroneous personal codes or without a fixed address, persons who had participated in a previous household survey conducted by Statistics Estonia during 2011 or 2012
Ireland	Non-private households, homeless
Greece	Population in institutions, homeless
Spain	Population in institutions, homeless
France	People who do not live in a main residence, people in institutions, homeless
Italy	Population in institutions, homeless, individuals not in the population register
Cyprus	Population in institutions, homeless, population of the areas of the Republic of Cyprus not under the effective control of the Government of the Republic of Cyprus
Latvia	Population in institutions, homeless, population in collective households
Luxembourg	Population in institutions, homeless, international civil servants and individuals not registered in the social security register in general
Hungary	Population in institutions, homeless
Malta	Population in institutions, homeless, immigrants living in assigned accommodation
Netherlands	Population in institutions, homeless, blind people, people who do not speak Dutch
Austria	Population in institutions, homeless
Poland	Non-private households, homeless
Portugal	Population in institutions, homeless, population living in collective dwellings
Slovenia	Population in institutions, homeless , people who do not report their current main residence to authorities
Slovakia	Collective households, homeless
Finland	Population in institutions, homeless

Note: Population in institutions refers to persons living in e.g. homes for elderly people, military compounds, prisons and boarding schools.

4.2.4 Use of replacements

A replacement of a sample unit occurs when a non-responding unit is replaced by another reserve unit during the fieldwork. Using replacements may help draw information in particular from groups of households that are most difficult to reach. On the other hand, replacements may have different characteristics from those of non-respondents and using replacements may result in a reduction of interviewers' efforts to get an interview from the originally selected unit. In the HFCS, the use of replacements is subject to strict control. Replacements are selected to closely match the replaced units in terms of important characteristics, and replacements are allowed only after special efforts have been made to convert refusals.

Replacements were used in three countries. In Slovenia, it was possible to use replacements in the first HFCS wave, but not in the second. Although the rules for using replacements varied, all countries followed the criteria mentioned above to a large extent.

In Spain, tightly controlled replacements were chosen. In large cities and provincial capitals, up to four replacements were provided for each original household in the

sample that would serve as replacements for that household only. These replacements were the two households immediately before and the two immediately after the household in a list ranked by income quartile (for non-filers of wealth tax), wealth stratum, and per capita household income. Replacements had to belong to the same income quartile (for non-filers of wealth tax returns) or the same wealth stratum as the sample household. In the case of smaller municipalities, Navarre and the Basque country, a more standard scheme of a pool of eight replacement households as potential substitutes for eight sample households within the same primary sampling unit was adopted.

In Italy, replacements are allowed within the same municipality after three unsuccessful contacts, on different days and at different times, determining not-at-home, refusals or ineligibility.

In Cyprus, replacements were selected from the same stratum as the original sample unit.

4.2.5 Oversampling of the wealthy

In wealth surveys, there are several additional challenges for the sample design in comparison to other household surveys. On the one hand, wealth surveys usually aim to conduct several kinds of analyses on all parts of the distribution. The previous parts of this chapter provide an assessment on how well inferences can be drawn from most parts of the wealth distribution. On the other hand, it is known that the distribution of wealth is skewed, and some types of assets are possessed only by a small fraction of households. Consequently, for the sample to adequately represent the full distribution of wealth in the population, it is essential to have a relatively high proportion of wealthy households in the sample (Kennickell, 2007). Data on the wealthiest households should be collected as efficiently as possible to get unbiased estimates of total wealth.

Furthermore, the general picture of wealth inequality will be negatively affected by the inability to collect data from the top fractions of the distribution. This will have an impact on indicators such as the Gini index, the share of wealth owned by the top 1%, and quantile ratios (for example, the ratio of net wealth between the households in the top 20% and bottom 20% of the wealth distribution), which are sensitive to the values of the richest households. Recently there have been attempts to measure the bias caused by the inability of survey data to sample the wealthiest households in the population with the help of external sources, such as data from Forbes' *The World's Billionaires* list (Vermeulen, 2014).

Capturing the values of assets from the wealthiest households is even more relevant in the case of certain individual items, particularly financial assets that are owned only by a small share of households.

In addition, there is evidence from previous wealth surveys that unit non-response rates are higher for wealthier households. This is first caused by the special difficulty of establishing contact with wealthy respondents, since they are more likely to be

absent from their principal residence during prolonged periods of time, to possess more than one residence and to be surrounded by additional security measures. In addition, both available time and self-perceived value/time ratios usually predispose wealthy households to refuse to take part in surveys.[20] If it is not compensated by post-survey adjustments, the different non-response rate would cause measurement bias. Furthermore, if the sample is selected using information correlated with wealth,[21] this same supporting information may also be useful in guiding post-survey adjustments, compensating for non-response and reducing sampling error.

In conclusion, a given level of precision would either require a rather large (and costly) sample or, if efficiently designed, a sample which should include a disproportionally high number of wealthy households. Indeed, using data from a purely random selection of units would thus yield a statistically very inefficient estimate of the distribution of wealth. These challenges should be anticipated during the sampling-design phase.

Fifteen out of twenty countries used different strategies to oversample wealthy households (Table 4.6). In addition, Slovenia oversampled regions with lower expected response rates. This is an improvement from the first wave, where nine out of 15 countries oversampled the wealthy. Compared with the first wave, Slovakia introduced oversampling strategies for the second wave. In addition, all new HFCS countries used oversampling.

The strategies varied significantly between countries, and were heavily dependent on the available data. Spain used individual data on taxable wealth, and France, individual data on net wealth. In Estonia, Finland, Latvia and Luxembourg, individual-level income, in Portugal, the size of dwelling and in Cyprus, household-level electricity consumption, were used as proxies for wealth. Other countries did not have access to personal-level income or wealth data, and consequently oversampling had to be based on regional-level information on income and/or property prices. Slovakia used a combination of regional-level income and personal-level labour status.

[20] For further information, see references in Sanchez-Muñoz (2011).

[21] For instance, register-based (such as on wealth or income taxes; property taxes; socio-economic information at municipality or small area level; census of dwellings; etc.) or survey-based information (either from previous waves of the survey or from other surveys).

Table 4.6

Oversampling strategies

Country	Criteria for oversampling	Details
Belgium	Regional average income and housing prices	Neyman allocation (a sample allocation method for stratified samples maximising survey precision with given sample size) based on income dispersion. Regional units with higher number of households and bigger dispersion of income are oversampled.
Germany	Regional indicators	Oversampling for high-income municipalities and wealthy street sections in municipalities with >100,000 inhabitants.
Estonia	Personal income	Oversampling rate based on having a sufficiently large subsample of wealthier households while retaining a nationally representative sample. Personal income data from tax registers.
Ireland	Deprivation/affluence indicator from census	Aim to have 20% of the overall sample from the top wealth decile.
Greece	Regional average income and real estate prices	In Athens and Thessaloniki, oversampling in areas where average income and real estate prices are in the top 10%.
Spain	Personal taxable wealth	Eight wealth strata based on taxable wealth, sample progressively larger in strata with higher taxable wealth, based on wealth and income tax returns.
France	Personal wealth data	Four strata oversampled: wealthy city dwellers, equity-based wealth, real estate-based wealth, lower wealth. Information derived from fiscal sources.
Italy	No oversampling	
Cyprus	Electricity consumption	61% of the gross sample was selected from households within the top 10% according to electricity consumption.
Latvia	Personal income	Different sampling fraction for highest income decile according to tax registers.
Luxembourg	Personal income	20% of the gross sample was drawn from the top income decile according to the social security register and the self-employed-headed fiscal household subpopulation.
Hungary	Average regional income	Localities with high average personal income oversampled.
Malta	No oversampling	
Netherlands	No oversampling	
Austria	No oversampling	
Poland	Regional income and property size	Four groups of wealthy households, based on income tax and size of properties. All these groups were oversampled to varying degrees.
Portugal	Dwelling size	50% of the sample drawn for dwellings with a floor space (m2) above a predefined threshold.
Slovenia	Regional	Ljubljana and Maribor, due to expected lower response rate.
Slovakia	List of high income streets, personal education and labour status	Tax office provided a list of streets with a high incidence of high income individuals (top 5% in the region) residents. Persons in those streets with labour status correlated with high wealth identified from census.
Finland	Personal income level and type	High-income earners and self-employed oversampled, based on personal taxable income of the main income earner of the household-dwelling unit. Data from tax registers and register of household-dwelling units.

Source: ECB – HFCS metadata.

The oversampling strategies have enriched the sample with a higher proportion of households with high asset values, or less common financial assets, leading to more precise estimates of wealth. However, the final representation of the wealthy in the sample is influenced by other factors, such as non-response. An indicator of the representation of the wealthy in the final sample is the “effective oversampling rate of the wealthy” (see Table 4.7). It indicates the extent to which the share of wealthy households in the sample exceeds their share in the population. These rates are given separately for households belonging to the richest 5% and 10% of the population.

To compute this indicator, the net wealth values of the 90th and 95th percentiles were first calculated from the weighted data. Subsequently, the (unweighted) shares of interviewed households exceeding these values were computed. When the net sample includes a relatively large number of wealthy households with small final estimation weights on average, it is an indication of high effective oversampling of the wealthy households.

Table 4.7

Effective oversampling rates of the wealthy

Country	Effective oversampling rate of the top 10%	Effective oversampling rate of the top 5%
Belgium	59	71
Germany	141	173
Estonia	31	35
Ireland	10	8
Greece	-2	-2
Spain	234	374
France	132	227
Italy	8	6
Cyprus	67	77
Latvia	53	62
Luxembourg	58	60
Hungary	2	-2
Malta	-4	1
Netherlands	54	67
Austria	-7	-12
Poland	10	10
Portugal	53	51
Slovenia	21	23
Slovakia	5	15
Finland	80	92

Notes: "Effective oversampling rate" of the top 10%: (S90 – 0.1)/0.1, where S90 is the share of sample households in the wealthiest 10%.
Effective oversampling rate of the top 5%": (S95 – 0.05)/0.05, where S95 is the share of sample households in the wealthiest 5%.
Wealthiest households are defined as having higher net wealth than 90% (95%) of all households, calculated from weighted data.

The interpretation of the figures in Table 4.7 is as follows: if the share of rich households in the net sample is exactly 10%, the effective oversampling rate of the top 10% is 0. If the share of households in the wealthiest decile is 20%, the effective oversampling rate is 100, meaning that there are 100% more wealthy households in the sample than there would be if all households had equal weights. A negative oversampling rate indicates that there are fewer wealthy households in the net sample than there would be if all households had equal weights.

A high effective oversampling rate means that the analyses of wealthy households – and accordingly of aggregate wealth and wealth inequality indicators – are more efficient. The range of oversampling rates is considerable in the HFCS. In the data for some countries, the share of wealthy households in the sample is smaller than their share in the population. In other cases, effective oversampling rates of the top

10% are up to over 200%, and the corresponding rates for the top 5% even higher. Judging by the previous table, oversampling strategies and data availability play a major role in the ability to get interviews from wealthy households. The availability of household-level information seems to be an especially big advantage.

In countries that participated in the first HFCS wave, oversampling rates are generally very similar or slightly higher in the second wave. Significant improvements in oversampling rates can be observed in Slovakia, where oversampling strategies were introduced in the second wave, and in Portugal.

5 Unit non-response and weighting

High unit non-response rates increase the variability of estimates drawn from the sample, and, to the extent that non-response is non-randomly distributed, it may lead to biased estimates of the variables of interest. Weight adjustments may to some extent be used to alleviate non-response bias.

This chapter compares indicators on response behaviour observed in the second wave of the HFCS and describes the common weighting procedure applied in the survey, along with the most significant country features on weighting and calibration. It also discusses an agenda for further related research.

5.1 Unit non-response in wealth surveys

Unit non-response is the failure to obtain information from an eligible sample unit. It is a result of either the inability to contact a selected sample unit, of the unwillingness of the sample unit to respond to the survey, or of several other reasons such as language barriers or inability to participate in the interview. Owing to the sensitivity of wealth data, observed unit non-response rates have been generally higher in wealth surveys than in income surveys.[22]

To improve the quality of the analysis to be conducted with survey data, it is generally considered essential that the basic survey weights determined by the sample design are adjusted to address non-response and other imperfections in the final sample, such as coverage problems. Furthermore, to maximise comparability in such a multi-national survey, it is usually seen as important that such procedures are common in each country, and are compatible with the structure of the sample and the data available for making adjustments.

Although a survey with a 20% response rate has a greater possibility for bias than a comparable survey with a 100% response rate, there is evidence that response rates and non-response bias are not always inversely related (Groves and Peytcheva, 2008). It is common practice to evaluate the degree to which there is identifiable response bias in a survey and the degree to which non-response adjustments may ameliorate such problems. In the case of the HFCS, it will also be important to investigate variations in national surveys that may lead to systematic differences in non-response bias.

[22] For further information, see references in Pérez-Duarte et al. (2010).

5.2 Unit non-response in the HFCS

The HFCS takes special care to minimise non-response rates to reduce non-response bias by emphasising the use of best practices. For example, emphasis has been put on interviewer selection and training, as well as on the incentives and workload the survey organisation offers to interviewers. To minimise variability in potential bias across the countries participating in the HFCS, emphasis is placed on the use of common practices, to the extent that this is feasible. Despite these efforts and the good flow of information and exchange of best practices across countries, there remained potentially important differences in procedures, such as the protocols used in directing attempted contacts with the survey respondents.

Table 5.1 presents indicators on response behaviour in the second wave of the HFCS. These indicators are based on standard definitions (see AAPOR, 2011). The following indicators are included:

- Response rate = Achieved interviews / Eligible sample units[23]
- Refusal rate = Sample units refusing to participate / Eligible sample units
- Cooperation rate = Achieved interviews / Contacted sample units
- Contact rate = Sample units contacted / Eligible sample units
- Eligibility rate = Eligible units / Gross sample size

The response rate is probably the most commonly used survey quality indicator. Because non-response reduces the number of observations available for analysis, it has direct implications on the sampling variability of survey estimates. Refusal, cooperation and contact rates provide useful information on the structural characteristics of non-response and may help to better administer survey resources towards respondents with a higher tendency to refuse participation in the survey, with a view to minimising the risk of non-response bias. Eligibility rates indicate the quality of the sampling frame.

There is a significant variation in the achieved response rates in the HFCS. In most cases, the main reason reported for unit non-response is refusal to participate, although contact rates are quite low in Latvia and Malta. In the comparison of response rates, it is worth noting that the Finnish figures refer to an income survey, and in France and Portugal, the survey is compulsory for households, though participation is never enforced. Moreover, in some countries, the HFCS was an adaptation of existing household surveys, and in seven countries, the survey also has a panel component. For countries with a panel component, both response rates of households interviewed for the first time and for the entire sample are given in Table 5.1.

[23] For sample units for which eligibility could not be defined during fieldwork, the share of eligible units is estimated from the corresponding share of those sample units for which eligibility was identified.

Compared with the first wave, response rates have increased quite substantially in Belgium (from 22% to 30%), Cyprus (from 31% to 60%) and Portugal (from 64% to 85%). Most other countries have experienced a decline in response rates, particularly Greece (from 47% to 41%) and Spain (from 40% to 31%).

Table 5.1

Response behaviour indicators in the HFCS

Country	Gross sample size	Net sample size	Response rate*	Response rate** (including panel)	Refusal rate	Cooperation rate	Contact rate	Eligibility rate
Belgium	7,265	2,238	30.0	38.4	45.2	40.8	94.2	80.2
Germany	16,221	4,461	19.0	29.0	57.0	31.0	93.0	94.4
Estonia	3,594	2,220	63.9		23.5	70.0	91.2	96.7
Ireland	10,522	5,419	59.7		36.9	66.6	89.7	86.2
Greece	7,368	3,003	40.8		50.6	41.7	97.7	100
Spain	13,442	6,106	31.7	48.1	46.0	49.6	97.0	94.4
France#	20,272	12,035	65.0		11.9	76.3	85.2	91.8
Italy	16,100	8,156	43.3	53.0	29.9	61.4	86.4	95.5
Cyprus	1,874	1,289	60.4	70.0	23.2	72.5	96.5	98.3
Latvia	2,405	1,202	52.9		27.4	65.1	81.2	94.5
Luxembourg	7,300	1,601	23.4		66.8	25.6	91.4	91.2
Hungary	17,985	6,207	38.5		14.8	40.3	95.6	89.9
Malta	2,035	999	35.4	51.0	30.3	61.7	82.7	96.2
Netherlands	2,562	1,284	32.0	50.1	49.9	50.1	100	100
Austria	6,308	2,997	49.8		44.1	51.2	97.3	95.5
Poland	7,000	3,483	54.2		32.1	55.6	97.5	91.8
Portugal#	8,000	6,207	84.8		5.1	92.2	91.9	91.5
Slovenia	6,519	2,553	40.5		41.8	46.7	86.7	96.6
Slovakia	4,202	2,136	53.4		32.0	61.8	86.3	95.2
Finland	13,960	11,030	64.1	80.1	13.7	83.5	95.9	98.6

Source: ECB – HFCS metadata.
Gross sample includes panel households that have responded to previous waves of the same survey.
In France and Portugal, survey participation is compulsory for households.
* For comparability, response rates are shown for households interviewed for the first time.
** Response rates for the whole sample in countries that have a panel component. In Finland, the panel component consists of households interviewed in the three previous waves of the income and living conditions survey.

Finally, it is worth mentioning that oversampling of wealthy households may lead to diminished response rates. In spite of this possible drawback, oversampling of specific population groups is beneficial for survey quality, and should be noted when comparing the response rates of individual surveys.

5.3 Weighting

Weighting procedures are an essential tool for adjusting, to the degree that this is possible, both for the bias caused by unit non-response and for other irregularities in the sample. In the HFCS, all participating surveys follow common high-level weighting procedures to ensure the comparability of survey data. There are minor differences in some of the details of implementation across countries participating in

the HFCS. In addition, there are differences in more granular elements, such as the structure of the samples and the frame-based and external sources used to adjust the weights.

5.3.1 Weighting procedures in the HFCS

The standard HFCS procedure for computing and adjusting survey weights takes into account: (i) the unit's probability of selection; (ii) coverage issues; (iii) unit non-response; and (iv) an adjustment of weights to external data (calibration). The methodology is coherent with existing international standards (Eurostat, 2011a and United Nations, 2005). These steps are implemented sequentially as follows:

Design weights are computed as the inverse of the selection probability of each unit in the gross sample, that is, both responding and non-responding units.

The first-stage weights are adjusted for coverage, including adjustments both for non-eligible units in the gross sample (frame over-coverage) and for multiple selection probabilities. This stage of adjustment is relevant especially for sampling frames designed from registers of dwellings rather than of households or individuals.

The coverage-adjusted weights are further adjusted in an attempt to minimise bias potentially induced by discrepancies between characteristics of survey respondents and non-respondents. This adjustment involves estimating response probabilities as functions of characteristics available for both responding and non-responding households, and dividing the coverage-adjusted weights of each responding unit in the achieved sample by the response probability. Such adjustments can be specific to individual units, but in the HFCS adjustments, they are made at the group level.

To obtain final weights, the non-response-adjusted weights are modified using auxiliary information to align the estimates of a set of variables with corresponding population estimate totals and category frequencies (Särndal, 2007). This adjustment of weights is motivated by a desire to reduce bias induced by discrepancies between the initial sample and the total population that are not captured in the coverage adjustments or that are induced through the other stages of weight adjustment. The HFCS uses a methodology that adjusts weights so that their totals by groups match their representation in the full population of households. To be effective, the calibration variables must be strictly comparable in both the survey and the source of the population data, correlated with the study variables, but not too closely correlated with each other. While the selection of calibration variables varies by country, partly dependent on available data sources, calibrating for at least age, gender and household size is common across all countries in the HFCS.

In surveys that have a panel component, the weighting procedure includes additional features, for which the HFCS has provided detailed guidelines. First of all, personal – and ultimately household – weights need to be adjusted for persons leaving and entering the households between waves. Secondly, household weights need to be adjusted for attrition and for households leaving and entering the target population. Different survey waves are treated as independent samples in the first stage of the

weighting procedure, and subsequently the samples are merged and their weights adjusted to the target population of the current wave before the final calibration step. With the exception of the Netherlands, where the panel component is not taken into account in the construction of weights, the guidelines on panel weighting are followed in all countries with a panel component in the HFCS data, as well as in Finland. Although the Finnish HFCS does not have a panel, the Finnish sample consists of four rotational groups of the income distribution survey, which are weighted separately, and finally panel-specific cross-sectional weights rescaled in proportion to the sample share of each group.

In sample surveys where different units have unequal probabilities of being sampled, using the inverse selection probabilities in weight construction will produce unbiased estimates of means and totals (Horvitz and Thompson, 1952). However, the variability of weights often increases the sampling variances of important survey estimates relative to those of a sample of the same size without weight variation, and there is a trade-off between unbiasedness and the efficiency (low variance) of estimates (Little, 1991). In the case of highly variable weights, the efficiency of estimates can be increased by trimming extreme weights.

Extreme final estimation weights were only trimmed in the surveys carried out in Italy, Malta and Finland (whereas in Germany, extreme weights were trimmed before the final calibration step). In the calibration, limits for weight adjustment factors can be set in order to define a ceiling for the ratio between design weights (adjusted for coverage or non-response) and final weights. This procedure was applied in Belgium, Germany, Spain, Hungary, Poland and Finland.

5.3.2 Variables used for calibration

Table 5.2 indicates the external variables and sources used in calibration. Note that in some cases, combinations of individual variables (for example, age by region or by municipality size) were used.

Table 5.2

Calibration variables and sources

Country	Age	Gender	Household size	Region	Other	Source
Belgium	X	X	X	X		Population statistics (NSI)
Germany	X	X	X	X	Municipality size, home ownership, size of main residence (for homeowners); education, labour status and nationality	Micro census
Estonia	X	X	X	X	Degree of urbanisation, education, ethnicity, home ownership	Census, EU-SILC
Ireland	X	X	X	X	Home ownership, deprivation	Quarterly national household survey
Greece			X	X	Home ownership	EU-SILC
Spain	X	X	X		Municipality size	Census
France	X	X		X	Degree of urbanisation, education and socio-economic status of reference person, household type, labour and wealth income	Census. LFS
Italy	X	X		X	Municipality size, income and labour status for panel households	Census
Cyprus	X	X	X	X	-	Census
Latvia	X	X		X	Income	Population statistics, tax register
Luxembourg	X	X	X		Nationality, labour status	Social security register
Hungary	X	X		X	Labour status, type of locality	Census. LFS
Malta	X	X	X	X	Labour status	NSI, LFS
Netherlands	X	X			Home ownership, education	EU-SILC, NSI
Austria*			X	X	Home ownership	Micro census
Poland	X	X	X		Urban/rural area	Census
Portugal	X	X	X	X	Loans for house purchase	Population statistics, LFS, Credit register
Slovenia	X	X	X	X	-	Population statistics
Slovakia	X	X	X	X	Labour status	Census
Finland	X	X	X	X	Type of municipality, selected income variables + number of income recipients, mortgage interest repayments, value and recipients of mutual funds, number of persons having listed shares	Population information system, tax and other income registers, register file on the values of listed shares

LFS: Labour force survey. NSI: national statistical institute. EU-SILC: EU Statistics on Income and Living Conditions. CBS: Central Bureau of Statistics, Netherlands.
* Cell-based post-stratification

5.3.3 Weights

The outcomes of the weighting procedures are shown in Table 5.3, including the sums, means and coefficients of variation of final estimation weights by country. The sum of final estimation weights corresponds to the size of the target population, i.e. the number of households. Mean weights indicate the average number of households that one net sample unit represents.

Table 5.3

Final estimation weights by country

Country	Sum	Mean	Coefficient of variation, %
Belgium	4,796,647	2,143	106
Germany	39,672,000	8,896	118
Estonia	571,857	258	49
Ireland	1,690,073	312	52
Greece	4,266,745	1,421	55
Spain	17,429,812	2,855	148
France	29,017,678	2,411	85
Italy	24,694,122	3,028	89
Cyprus	303,242	235	156
Latvia	828,907	690	87
Luxembourg	210,965	132	70
Hungary	4,127,671	665	64
Malta	159,427	160	63
Netherlands	7,590,228	5,911	56
Austria	3,862,526	1,289	41
Poland	13,492,882	3,905	60
Portugal	4,017,981	647	114
Slovenia	820,541	321	66
Slovakia	1,855,392	869	104
Finland	2,622,499	238	81

Notes: Sum is the sum of the estimation weights over the households, and corresponds to the size of the target population, i.e. the number of households. Mean weights indicate the average number of households that one net sample unit represents. The coefficient of variation is the relative standard deviation of final estimation weights (as a percentage of the mean of weights). This indicates the variability of the final weights in the net sample.

6 Editing, item non-response and multiple imputation

Data editing is an essential part of processing survey data in order to minimise the errors and inconsistencies from collected observations. The first part of this chapter describes the editing process in the HFCS, and provides information on the share of edited observations in various countries.

In any household survey, a certain degree of item non-response is always expected. In a wealth survey like the HFCS, which contains difficult and sensitive questions on personal finances, one can expect a higher level of missing answers, and in particular for some of the most important variables used in the production of statistical indicators and as components of research models. Imputation is the most frequently used process of correcting for item non-response by assigning plausible values to a variable when it was not collected at all or not correctly collected based on the information collected from other households.

6.1 Data editing

To obtain accurate survey results, data must be, to the greatest extent possible, free from errors and inconsistencies, especially after the data processing stage. The procedure for detecting errors in and between data records, during and after data collection and capture, and for adjusting individual items is known as editing (UN, 2001). Editing is a critical step in maintaining data quality. Kennickell (2006) shows the effect of editing the data in the Survey of Consumer Finances by comparing the distributions of net worth of imputed but unedited data with the imputed and edited data. The unedited data show, for example, underestimation at the bottom of the distribution, but strong overestimation at the top. The Gini index on net worth is significantly higher in the unedited data.

The use of carefully programmed computer assisted interviews can significantly reduce the number of consistency checks needed after the fieldwork phase. Furthermore, comments made by interviewers during data collection can help in identifying possibly unreliable values (Bledsoe and Fries, 2002). In all countries conducting the HFCS, consistency and range checks were included in the questionnaires. In most cases, interviewer comments were used systematically in the review of data values.

As a first option in editing values that do not seem coherent, interviewers can re-contact households to verify values of individual variables. This procedure was possible in most HFCS countries. However, information on the number of households re-contacted is available only in individual cases (see Table 6.1).

Table 6.2 shows the shares of edited observations for the value of the household main residence and the value of savings accounts, as well as the number of

variables with relatively high edit rates. For most variables, the shares of edited observations are very small. There are also several country-specific features in data collection that explain high edit rates in several cases and where high edit rates should not be interpreted as a result of low-quality data collection during interviews. Editing has been used for example to convert net amounts of income variables to gross amounts (see Table 9.4) or to complement interview information with administrative data (see Table 3.3).

Table 6.1

Information on data editing

Country	Organisation responsible for editing	Interviewer comments used in editing	Re-contacting of households possible*
Belgium	NCB	In most cases	No
Germany	NCB	Systematically	Yes
Estonia	NCB	Systematically	Yes
Ireland	NCB and NSI	Systematically	Yes
Greece	NCB	In most cases	Yes
Spain	NCB	Systematically	Yes
France	NSI	Sporadically	Yes
Italy	NCB and SA	Systematically	Yes
Cyprus	NCB	Systematically	Yes
Latvia	NCB	In most cases	No
Luxembourg	NCB	Systematically	Yes
Hungary	NCB and NSI	Sporadically	No
Malta	NCB	In most cases	Yes
Netherlands	SA	Not applicable	Yes
Austria	NCB	Systematically	Yes
Poland	NCB	No	No
Portugal	NCB and NSI	Systematically	Yes
Slovenia	NCB	In most cases	Yes
Slovakia	NCB	Sporadically	Yes
Finland	NSI	Systematically	No

* Only re-contacts for verification of data values included, re-contacting households for verification of data authenticity excluded.
Notes: NCB; National Central Bank, NSI: national statistical institute, SA: Survey Agency

Table 6.2
Edit rates

Country	Value of main residence (% of cases)	Savings accounts (% of cases)	Number of variables with edit rates >5%
Belgium	1.3	0.7	68
Germany	0.2	0.0	14
Estonia	0.6	0.0	32
Ireland	1.0	0.0	4
Greece	0.0	0.0	19
Spain	0.0	0.0	0
France	0.0	0.0	0
Italy	0.7	0.0	5
Cyprus	0.4	0.3	3
Latvia	26.2	0.0	120
Luxembourg	0.0	0.0	50
Hungary	0.0	33.6	3
Malta	10.1	2.6	36
Netherlands	0.0	0.0	3
Austria	0.0	0.1	4
Poland	5.6	-	9
Portugal	0.3	0.3	7
Slovenia	0.1	0.1	3
Slovakia	0.2	0.0	5
Finland	0.0	0.1	0

Source: HFCS

6.2 Imputation of the HFCS data

In the HFCS, observations for which no valid response was received from the households should be imputed. In addition to a common methodology on imputations, software tools have been developed for imputation in order to maximise the degree of methodological commonality.

6.2.1 Basic common rules

A complete-case analysis that discards non-observed units and analyses only units with complete data would disregard too much information and is thus not considered appropriate for the HFCS. Inferences should be made from the survey data on the entire population rather than on only those units that have provided answers to certain questions (Little and Rubin, 2002). While a requirement to impute all missing values for all variables was not realistic for the first HFCS waves, a minimum set of variables that need to be imputed was determined for the first wave (Household Finance and Consumption Network, 2008b).[24] The set of 130 variables that were

[24] See Biancotti et al. (2009) for additional references.

fully imputed in the first wave basically included all components of household income, consumption and wealth, so that the indicators on households' balance sheets could be based on the observations of all households that participated in the survey.

For the second wave, this minimum list of variables to be imputed was updated with 57 new variables. The updated list now includes new balance sheet variables on private loans, as well as selected variables that are most frequently used in the reporting of HFCS results, in monetary policy and financial stability analysis, and as good predictors of balance sheet variables in the imputation models.

The need to provide information about the quality of the data to the users is recognised. For this purpose, a set of shadow variables (so-called flag variables) is produced and provided to users to indicate the origin of the information corresponding to all variables and observations. Flag variables indicate, for example, whether an individual observation was recorded as collected, edited, estimated, imputed from a range value provided by the respondent, or imputed because the respondent could not or did not want provide a valid response.

Each NCB/NSI that produces the data has the responsibility to impute missing observations. Rubin (1996) makes the case explicitly, claiming that modelling the missing data must be, in general, the data constructor's responsibility, since "in general, ultimate users have neither the knowledge nor the tools to address missing data problems satisfactorily." Database constructors using individual HFCS country data have better information on the reasons for non-response and on the relationship between different variables. Besides, country-specific questions or different interviewing strategies are better evaluated at the country level. Finally, part of the information used in the construction of the imputation models is only available at the country level due to confidentiality reasons (wealth strata, regional data, interviewer comments and so on). Against this background, although the HFCS imputation process strictly follows a common methodology (see next sections), its implementation is fully decentralised at the country level.

6.2.2 Multiple imputation

The goal of imputation is to preserve the characteristics of the distribution of and the relationships between different variables (Rubin, 1987). In addition to a complete-case analysis, several other simple procedures could be performed to deal with missing values.

Probably the simplest approach is to fill in missing values with the means of observed values. This would naturally lead to a large decrease in variance and would not reproduce the distributions obtained from the survey data. In stochastic regression imputation, missing values are replaced with a value predicted by a regression plus a residual, to reflect the uncertainty in the predicted value. For normal linear regression models, the residual is normal, with zero mean and variance equal to the residual variance in the regression. For binary or multinomial regressions, the predicted value is a probability distribution and the imputed value is

drawn from that distribution. While this method preserves the distribution of the imputed values, the uncertainty of the imputation process is not fully reflected in a single imputation.[25]

With multiple imputation (MI), M imputed values based on different random draws are provided to the user for each missing value, resulting in M copies of the complete dataset. MI shares the advantages of single imputation in that it allows complete-data methods of analysis and use all the information available to the data collector. However, with MI, uncertainty can be taken into account (i.e. in order to avoid underestimating the resulting variance), which is particularly important in cases of significant item non-response.

The construction of multiple imputation models in the HFCS is based on the methodologies used in similar surveys by the Federal Reserve Board and Banco de España (see Kennickell, 1991 and 1998, and Barceló, 2006). HFCS datasets include five implicates (imputed sets of values) for each missing observation. The distance between the five implicates accounts for the underlying level of uncertainty. The imputation technique has an iterative and sequential structure. The models follow a path in which all variables are filled in with a predefined sequence. The models are run iteratively several times, and imputed values from each of the previous rounds are treated as observed values in the subsequent iterations.

Furthermore, a broad-conditioning approach is used, meaning that a high number of covariates, based on several criteria, are included in the models for all variables to be imputed. The model should include, first of all, variables that have predictive power, empirically shown by regressions, for the variable to be imputed. Covariates should also include variables that have explanatory power suggested by economic theory, although not empirically exhibited for the dataset in question. Because of the sequential structure of the model, predictors of the most frequently used covariates for other variables are also important. Finally, any variables that could potentially explain the non-response pattern of households should appear as covariates in the imputation model. MI in the HCFS is based on the assumption of “missing at random”, meaning that the distribution of the complete data only depends on the observed data, conditional on the determinants of item non-response and other covariates. Consequently, this complete set of variables has to be incorporated to the imputation models (Barceló, 2006).

6.2.3 Methodology and common software tools

In multinational surveys, countries should use similar methodologies to impute missing values. While the exact structure of models is always country- and data-dependent, using the same or at least similar methodological tools preserves data comparability. To maximise the degree of methodological commonality, the HFCS has cooperated in the development of common software tools for imputation.

[25] For further information, see references in Household Finance and Consumption Network (2008b).

A common SAS software package called €MIR has been developed for the purpose of multiply imputing HFCS data. This set of programs produces diagnostics on the missing values and an overview of descriptive statistics, prepares the data for imputation and analyses imputed results. The main part of the program, the imputation model itself, is based on the FRITZ program created for the imputation of the Survey on Consumer Finances at the Federal Reserve Board. The program is structured as an SAS macro embedded in a wider framework determined by the implementation of Gibbs sampling. Gibbs sampling is an iterative Markov procedure of successive simulation of the distribution of variables conditioned on both observed data and distributions of variables previously simulated in the same iteration. The model imputes each missing observation using a maximal set of covariates (from the list determined by the user) from the appropriate subpopulation. For example, in the imputation of the value of bonds, only households that have bonds are considered (Kennickell, 1991).

Common imputation tools have also been developed for the Stata software. The imputation model in a software package called ICE (Royston, 2004) is based on the same multiple imputation algorithm and implementation of Gibbs sampling as €MIR. While there are some minor differences in dealing with certain types of observations (i.e. using pooled samples in the case of similar variables, such as different loops for the same item or imputing variables reported in ranges), few differences should be expected in the outcome of the same imputation models in comparison to €MIR.

A Stata software package called MeDaMi was developed in the network. In MeDaMi, the specification of suitable imputation models is automated, and the user only needs to revise and verify the set of covariates used in the models prior to executing the imputation procedure. While the method of automated determination of covariates allows for a significant reduction in human resources, it might diminish the data producer's incentives to fully examine the relationships between different variables, missingness patterns, etc. that are vital in the construction of good quality imputation models.

Of the 20 countries participating in the second wave of the HFCS, 16 used MI to correct for item non-response.[26] The exceptions were France, Italy, Ireland and Finland. In Italy and Finland, the level of item non-response was very low for different reasons. In Italy, the low level of item non-response was due to the specificities of the contract with the survey company.[27] Consequently, single imputation was used, and the imputed values result from a regression model with a random component. In Finland, most balance sheet and income variables are register data or produced using register-based estimation, and the share of missing information for most variables that were collected was negligible.

Descriptions of some of the most important methodological choices for the imputation models are presented in Tables 6.3 and 6.4. In Table 6.3, the first column

[26] Hungary used multiple imputation for 33 variables and single imputation for the remaining imputed variables.

[27] The contract with the survey company only considers interviews with a level of item non-response below a certain threshold as completed cases.

shows whether survey weights are used in the imputation models – either by performing weighted regressions or by using survey weights as covariates. There is evidence that ignoring information on sampling design in the imputation models will lead to biased results (Reiter et al., 2006; Zhang et al., 2009). However, weighted regression potentially leads to less efficient estimates (Faiella, 2010).

The second issue in Table 6.3 indicates whether limits were introduced for the number of collected observations, below which missing values were not imputed for a variable (apart from the natural limit of two observations, below which imputation is not technically possible). A low number of collected observations will naturally add uncertainty to the imputation model. One way to solve this problem is to pool several variables to achieve a sufficient number of observations (for example, merging several loops of one type of mortgage).

The last item in Table 6.3 describes the selection process of covariates for the imputation model. The automatic model specification with limited editing is a feature of the Stata/MeDaMi described earlier, and was used in four countries. Other countries evaluated the selection of covariates on a case-by-case basis, with some kind of automatic pre-selection process in some cases.

Table 6.3

Imputation methodology

Country	Use of weights			Limiting imputation due to low number of observations				Selection of predictors in the imputation model		
	Weighted regression	Weight as covariate	No weights used	Yes, values left missing	Yes, estimation/ other methodologies used	Yes, variables pooled and imputed	No	Automatic with limited editing	Automatic pre-selection with case-by-case evaluation	Case-by-case evaluation
Belgium	X					X			X	
Germany	X						X		X	
Estonia			X	X						X
Ireland			X	X						X
Greece		X				X				X
Spain			X	X						X
France			X				X			X
Italy	X						X			X
Cyprus			X		X			X		
Latvia	X	X*					X		X	
Luxembourg		X					X		X	
Hungary	X				X					X
Malta		X			X			X		
Netherlands			X	X	X				X	
Austria		X					X			X
Poland	X						X			X
Portugal	X			X						X
Slovenia			X		X			X		
Slovakia		X			X			X		
Finland	X						X		X	

Source: ECB – HFCS metadata.
*In individual cases

Table 6.4 shows the numbers of covariates used in the models to impute four of the most significant balance sheet and income variables: the current value of the household main residence, the outstanding balance of the biggest loan collateralised by the household main residence, the value of savings accounts and employee income. These figures indicate significant differences in the degree to which the broad conditioning approach (see Section 6.2.2) was applied in various countries. The use of a large set of covariates in the imputation models is recommended to preserve the association between different variables in the dataset.

These figures are not perfectly comparable, since there was a large variation in the numbers of variables collected in different countries, as well as in the sample sizes. For example, regional data is collected for national purposes only in some countries. In countries that collect these data, numerous dummy variables are often created from regional variables to be used as covariates in the imputation models. In addition, there might be several imputation models for one variable, e.g. if a variable is an aggregation of several variables or if different models are constructed for different subsamples. The figures shown in Table 6.4 indicate the maximum number of covariates used in the imputation models for each output variable.

Table 6.4

Number of covariates used for main variables

Country	Value of Household Main Residence (HMR)	Outstanding amount of most important HMR loan	Value of savings accounts	Employee income
Belgium	116	108	118	70
Germany	83	7	17	20
Estonia	12	6	2	4
Ireland	3	2	2	3
Greece	70	77	71	79
Spain	241	111	162	251
France	24	14	20	n.a.
Italy	4	4	4	4
Cyprus	102	31	69	72
Latvia	50	50	23	35
Luxembourg	49	10	43	47
Hungary	2	1	1	4
Malta	14	16	14	58
Netherlands	17	15	18	12
Austria	153	128	51	72
Poland	6	4	7	6
Portugal	19	14	22	18
Slovenia	37	6	25	38
Slovakia	46	28	26	81
Finland	n.a.	n.a.	14	n.a.

Source: ECB – HFCS metadata.

6.3 Comparative information on item non-response and imputation

This section presents data on the outcome of the imputation process for all 20 countries that participated in the second wave of the HFCS. This section looks at the level of item non-response for the most important variables. These indicators reflect the degree and quality of imputations in different countries.

6.3.1 Item non-response rates for main variables

Tables 6.5-6.7 show information on the imputed observations for three of the most significant balance sheet variables: the current value of the household main residence, the outstanding balance of the biggest loan collateralised by the household main residence and the value of savings accounts. The first two columns indicate the share of households or persons at least 16 years old that have either reported having the item or for which the item was imputed as existing. The next three columns show the share of non-missing observations that were collected, imputed from a range value provided by the respondent or imputed from a missing value, respectively. The last two columns show the difference between the conditional means of all and collected observations.[28]

With very few exceptions, the variables indicating the existence of the items mentioned above were collected in the interviews. However, the share of imputed values for the values of these items is sometimes relatively high, and the imputation rates vary between countries and variables. In some countries, particularly in Malta, Austria and Portugal, a high share of balance sheet values has been imputed from a range value provided by the respondent. This procedure should be distinguished from an imputation for a missing value, since the range value provides a fair estimation of the point value directly received from the respondent.

The value of the household's main residence turned out to be the easiest one to provide for the respondents, with imputation rates remaining below 10% in most countries. Values of outstanding loan balances and savings accounts were clearly more difficult to collect, and a high variability in the imputation rates between various countries can also be seen.

The mean values of individual items do not, in most cases, change notably when imputed values are disregarded. This is somewhat to be expected, given the low share of imputed values. In individual cases, the imputed values of some variables have a significantly higher or lower mean compared with the collected values. This should indicate that households that were not able to record these items are expected to have higher or lower values for the corresponding variables than average households, given the covariates used in the imputation model. A large

[28] As has already been mentioned, in Finland these items are collected directly from registers or via register-based estimation, while in Italy the features of the contract with the survey company has produced extremely low item non-response rates.

difference between the imputed and collected values does not necessarily imply a biased imputation, it may just be a reflection of the differences between households that are able to provide asset values in the interview and households that are not.

In the comparison of item non-response rates, a few issues should be noted. As mentioned in the previous chapter, the surveys in France and Portugal are compulsory. While this has a positive impact on the response rates, it could have a detrimental impact on the motivation of respondents to provide all information needed, and hence increase item non-response. In some countries, particularly in those adapting the HFCS to an existing survey and to some extent also in Germany, the HFCS blueprint questionnaire was not implemented as such. A part of the HFCS variables were converted from variables collected in more detail for national-level purposes. Interviewing in more detail, as well as differences in the routing of the questionnaire, might overstate item non-response in the HFCS data compared with national data. When one HFCS variable is constructed from several national variables, non-response to any of the involved national questions is reflected in the HFCS variable.

Table **6.5**

Item non-response rates: current value of household main residence

	% having item		Of those having item*			Conditional mean (EUR)	
Country	Reported having item	Imputed as having item	Collected	Imputed from ranges	Imputed from missing	All	Collected#
Belgium	72.0	0.2	86.8	9.1	2.8	272,200	272,900
Germany	58.5	0.2	93.7	3.8	2.3	236,700	239,000
Estonia	80.1	0.0	69.8	0.0	29.6	71,000	71,700
Ireland	70.2	0.0	93.5	0.0	5.5	191,900	193,400
Greece	68.6	0.1	74.0	13.8	12.3	83,100	83,600
Spain	87.5	0.0	87.5	8.8	3.8	189,200	189,900
France	70.3	0.1	12.4	66.6	21.0	220,100	236,800
Italy	72.0	0.0	99.3	0.0	0.0	218,500	218,500
Cyprus	79.1	0.0	70.5	0.0	29.1	231,700	225,100
Latvia	77.4	3.2	68.7	2.3	2.8	31,500	31,800
Luxembourg	73.4	0.0	87.5	7.8	4.7	654,600	653,700
Hungary	84.8	0.0	99.5	0.0	0.5	38,600	38,600
Malta	78.5	0.0	37.0	52.8	0.1	213,600	213,600
Netherlands	70.8	0.0	90.2	0.0	9.8	244,000	246,600
Austria	42.8	0.0	76.5	19.9	3.6	291,500	292,300
Poland	77.5	0.0	67.2	15.2	12.0	89,700	86,800
Portugal	81.3	0.1	73.1	19.1	7.5	109,700	110,700
Slovenia	76.5	0.1	86.3	0.0	13.7	110,400	112,200
Slovakia	86.7	0.0	90.1	5.1	4.5	61,500	60,900
Finland	77.3	0.0	All values estimated			191,400	191,400

* Collected observations include those collected from administrative sources. In addition to collected and imputed values, observations can be edited or estimated, which is why the columns do not always add up to 100%.
\# Includes observations collected from registers, edited, estimated or collected as range values and then imputed.

Table 6.6

Item non-response rates: largest mortgage on household main residence: value still owed

Country	% having item		Of those having item*			Conditional mean (EUR)	
	Reported having item	Imputed as having item	Collected	Imputed from ranges	Imputed from missing	All	Collected#
Belgium	25.1	0.6	73.9	16.6	8.0	84,100	85,100
Germany	22.5	0.6	93.4	2.9	3.4	81,900	84,000
Estonia	22.9	0.0	94.2	0.0	2.9	38,400	39,200
Ireland	34.4	0.0	78.7	0.0	16.0	143,800	150,500
Greece	10.4	0.0	69.2	0.0	25.0	43,900	48,500
Spain	19.0	0.1	94.2	2.7	3.1	79,800	80,100
France	22.5	0.0	81.0	0.0	19.0	89,200	92,000
Italy	7.9	0.0	88.2	0.0	0.0	75,600	75,600
Cyprus	41.4	0.0	87.8	0.0	12.0	121,500	130,800
Latvia	13.7	4.0	15.5	0.0	0.5	33,500	33,500
Luxembourg	33.9	0.3	85.3	6.0	8.4	201,300	196,500
Hungary	17.7	0.0	98.2	0.0	1.8	15,300	15,200
Malta	12.2	0.0	67.2	32.8	0.0	63,100	63,100
Netherlands	49.2	1.1	85.2	0.0	14.8	125,700	129,200
Austria	12.5	0.5	60.4	14.8	24.8	83,100	78,500
Poland	10.8	0.0	71.4	0.0	20.1	33,100	34,800
Portugal	35.6	0.1	74.4	18.8	6.6	68,700	68,700
Slovenia	8.7	0.1	81.4	0.0	18.6	44,800	46,900
Slovakia	11.4	0.1	66.6	8.2	20.7	25,200	26,400
Finland	38.6	0.0	100	0.0	0.0	77,700	77,700

* Collected observations include those collected from administrative sources. In addition to collected and imputed values, observations can be edited or estimated, which is why the columns do not always add up to 100%.
# Includes observations collected from registers, edited, estimated or collected as range values and then imputed.

Table 6.7

Item non-response rates: value of savings accounts

Country	% having item		Of those having item*			Conditional mean (EUR)	
	Reported having item	Imputed as having item	Collected	Imputed from ranges	Imputed from missing	All	Collected#
Belgium	77.6	0.4	81.1	12.2	6.0	40,500	41,200
Germany	79.2	0.2	90.9	4.4	4.6	29,300	29,400
Estonia	39.8	0.0	100	0.0	0.0	9,200	9,200
Ireland	52.3	3.9	87.4	2.7	9.9	29,900	29,500
Greece	60.6	0.8	71.9	8.0	20.1	9,500	7,100
Spain	30.9	0.6	85.5	6.2	8.3	44,900	44,800
France	89.2	0.2	69.4	23.8	6.8	17,600	17,900
Italy	25.9	0.0	62.9	37.1	0.0	13,600	13,600
Cyprus	30.3	0.2	84.1	0.0	15.6	39,300	40,900
Latvia	12.8	0.1	72.3	3.2	24.5	7,000	6,900
Luxembourg	74.9	0.2	62.9	13.9	23.2	67,800	66,600
Hungary	50.2	0.0	97.8	0.0	2.2	5,800	5,900
Malta	86.9	0.6	29.7	55.0	12.6	25,000	25,900
Netherlands	88.2	0.3	98.7	0.0	1.3	23,500	23,500
Austria	82.9	1.9	67.9	17.9	14.1	26,900	25,400
Poland ##	82.1	0.0	58.9	27.6	13.5	3,800	3,300
Portugal	50.8	0.2	57.1	28.0	14.6	25,800	22,500
Slovenia	28.0	0.1	82.9	0.0	16.9	10,800	10,900
Slovakia	28.2	0.2	55.5	7.8	36.8	6,800	6,600
Finland	20.5	25.2	36.2	5.3	4.8	25,500	25,700

* Collected observations include those collected from administrative sources. In addition to collected and imputed values, observations can be edited or estimated, which is why the columns do not always add up to 100%.
# Includes observations collected from registers, edited, estimated or collected as range values and then imputed.
In Poland, savings accounts were not collected separately, all deposits included in sight accounts. The figures are related to sight accounts.

7 Variance estimation

Variance estimation is an essential element of survey data, as it allows researchers to distinguish between a statistically significant phenomenon and a spurious result caused by the random nature of the sample. Variance needs to be estimated, since the true value of the variance of an estimator can only be known if the values of the variables of interest in the whole population are observed. Underestimating the variance of an estimate may lead to incorrect conclusions (too many false positives), while overestimating the variance seemingly decreases the usefulness of the data, as fewer outcomes are estimated as being statistically significant.

Variance can have several components, though not all components can be estimated. One central component is the sampling error, which is caused by the random selection of the units participating in the survey. A second component is item non-response, which is addressed in Chapter 6 on Imputation, and which will be connected to total variance estimation in this chapter.[29]

Users of the HFCS need to be able to estimate the variance of several kinds of indicators. This chapter motivates the use of replication-based methods and describes the one chosen for the HFCS. The combination of replicate weights and multiple imputation is given in Section 7.3, and software routines for estimating total variance are sketched out in Section 7.5.

7.1 Motivation for replication-based methods

Since sampling error is linked to the sample design, its estimation relies on the provision of sample design information. In most surveys, the information on the number of stages of sampling, the strata at each stage, the identification of sampling units (primary, secondary, etc.) and the selection method (e.g. with or without replacement, equal or unequal probabilities) is sufficient to allow end-users to estimate sampling variance, using linearisation techniques for estimators other than means or totals. However, even in that case, with complex sample designs, these variance estimates are not simple to compute.

Moreover, sample design information is often withheld for confidentiality reasons: in many countries, the first level of stratification is often geographic (regions), and primary sample units are often linked to geographical units (municipalities, blocks, etc.). This increases the re-identification risk, and survey producers are understandably concerned about providing sample design information in that case.

[29] Other potentially relevant sources of variability, which the survey is not currently able to estimate, include variations in the understanding of questions by respondents, in interviewers' adherence to survey protocol, in formal sample coverage, and in decisions made in data editing or other aspects of processing.

Replication techniques are a robust and flexible way to estimate variance, even in the case of complex survey designs. Although in theory it applies only to linear statistics, and asymptotically in the case of the bootstrap, in practice these techniques have been found to be very useful because their flexibility allows them to cope with both different kinds of sampling designs and various kinds of statistics, without requiring an explicit formula for the variance of each statistic (as with linearisation techniques).

Nevertheless, the relative merits of different replication techniques are still under discussion (among them, Jackknife, Balanced Repeated Replication, and bootstrap, each with many variants). Replication techniques are similar in that in all cases, the full sample is used to draw (in different ways) sub-samples or replicate samples, which are used to estimate the statistic of interest and its variation across replicate samples, and which can be provided to users as a (large) set of replicate weights.

This chapter will not cover the different methods. Lehtonen and Pahkinen (2004) provide a good exposition and comparison of the different replication methods (called sample reuse methods in their book). We will focus hereafter on the bootstrap, as it was decided by the HFCN that the bootstrap offers the flexibility needed to cover the different national sample designs, and is powerful enough to cover many types of estimators.

In the bootstrap procedure, a with-replacement[30] sample of primary sampling units (PSUs) from each stratum is selected.[31] The number of PSUs per unit does not need to be constant. The number of replicates (bootstrap samples), as well as the number of PSUs sampled in each replicate, can be chosen by the analyst, although there are practical recommendations for both these quantities (for example, in the rescaling bootstrap proposed by Rao and Wu, 1988, and generalised by Rao et al., 1992). The precision of the bootstrap is higher if the number of replicates is increased.

Although the bootstrap has been slower to gain acceptance in the context of sample surveys, as it was originally developed for independent and identically distributed observations, improvements over the past 20 years have shown it to be a good alternative to other replication techniques (see Mach et al., 2007 for a description of its use in Statistics Canada, and Girard, 2009 for a general description).

7.2 The Rao-Wu rescaled bootstrap and its extensions

The variant of bootstrap for the HFCS is the rescaling bootstrap of Rao and Wu (1988), as further specified by Rao, Wu, and Yue (1992). It is applicable for one-stage samples, and can also be used in the case of a multi-stage sample drawn with low sampling fraction in the first stage. This is the case in several popular setups of stratified sampling. In addition, other sampling designs can be approximated by this

[30] Meaning each selection is independent, such that an element may be selected more than once and thus may appear multiple times in the same sample.

[31] In case of multi-stage sample designs, the methods below only consider the first sampling stage, as in practice this stage represents the largest part of the variance.

setup. While – like all bootstrap methods – the rescaling bootstrap is computationally intensive and the resulting variance estimates may be less stable than with other methods (such as Jackknife and linearisation), it provides consistent variance estimates in the case of non-smooth statistics such as distribution quantiles. Finally, the rescaling bootstrap has been implemented in SAS and Stata, and one of these two implementations has been used by all HFCN members.

The Rao-Wu bootstrap can be described as follows. We consider the case of strata indexed by $h = 1, \dots H$, with N_h units in each of them, out of which n_h are sampled without replacement. The sampling fraction is thus $f_h = n_h/N_h$. To each unit (h, i) there is a variable of interest y_{hi} and a weight $w_{hi} = N_h/n_h$. The total of this variable is $Y = \sum_{h=1}^{H} \sum_{i=1}^{N_h} y_{hi}$ which is estimated without bias by $\hat{Y} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi} y_{hi}$. The parameter of interest is a function of this total, say $\hat{\theta} = f(\hat{Y})$. For the Rao-Wu bootstrap applied in the HFCS, the following is done B times:

A sample of size m_h is taken with replacement from each stratum.

Writing r_{hi}^* the number of times unit (h, i) is resampled, the weights are adjusted as follows: $w_{hi}^* = \left(1 - \lambda_h + \lambda_h \frac{n_h}{m_h} r_{hi}^*\right) w_{hi}$ with $\lambda_h = \sqrt{\frac{m_h(1-f_h)}{n_h - 1}}$.

The bootstrap total is computed $\hat{Y}_b^* = \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi}^* y_{hi}$ and $\hat{\theta}_{*b} = f(\hat{Y}_b^*)$.

The bootstrap variance is then calculated as $V_{*(\theta)} = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}_{*b} - \bar{\hat{\theta}}_*\right)^2$, where $\bar{\hat{\theta}}_*$ is the mean of the bootstrap total over all B iterations.

7.2.1 Replicate sample size

In the HFCS, the replicate samples are drawn independently and with replacement in each stratum. The number of units m_h drawn in each stratum of size n_h are set to $m_h = n_h - 1$. The final estimation weight for each observation is then rescaled by a specific factor $\frac{n_h}{n_h - 1}$, and multiplied by the frequency of the observation in the replicate sample (number of hits).

7.2.2 Number of replicates

The number of replicates is at least 1,000, as a commonly used compromise between computational efficiency and stability of the variance estimates. Given the way bootstrap works, in practice it is not necessary to use all the weights. It is possible to only use e.g. the first 200 or 500 replicates for faster (but somewhat more unstable) variance estimation. This may depend on the type of estimator and size of the domain (e.g. mean of total population vs. medians for specific population subgroups). Some countries have provided more replicate weights (up to 2,000), in order to increase the stability of the bootstrap variance estimates.

7.2.3 Variance estimation model

Given that the standard Rao-Wu rescaled bootstrap is applicable to one-stage stratified simple random samples, and given the two- and three-stage designs used in some countries, a variance estimation model has been used in several countries. In particular, the second sampling stage is dropped (as in practice most of the variance originates from the first stage), except when the PSU is sampled with certainty, in which case the second sampling stage is used in the bootstrap. Strata may be merged, in particular if the number of units is small. In countries with dual-list samples, some adaptation of the methods was required.

7.2.4 Calibration of replicate weights

Since the final weights are adjusted for non-response (see Section 5.3 in Chapter 5 of this report), post-stratified or calibrated (the specific technique not being important), the replicate weights have been adjusted according to the same procedure, for example by running the calibration procedure with the same margins on each of the replicate weights. This can be considered an additional rescaling factor. For instance, after drawing the sample and rescaling the weights as in point 3, the weights are further rescaled to satisfy post-stratification or calibration constraints for each replicate. This is to ensure that the replicate estimates are close to unbiased in each replicate sample.

Table 7.1 shows information on the calibration of replicate weights. In most countries, each set of replicate weights sums up to the same number of households, consistent with the sum of final estimation weights (see Table 5.3), and to the same number of persons. When they do not, the variation of the number of households/persons is limited. Depending on the exact calibration used, there are some variations between each set of replicate weights in also in the population estimates by gender or age, indicated by the coefficients of variation in Table 7.1.

Table 7.1

Calibration of replicate weights and impact on population estimates

Country	At household level	At person level	By gender	By age group*
Belgium	Yes	Yes	Yes	(0.4%)
Germany	Yes	No (0.2%)	No (0.8%)	(1.3%)
Estonia	Yes	Yes	Yes	(0.6%)
Ireland	Yes	Yes	Yes	(0.8%)
Greece	Yes	No (0.4%)	No (1.0%)	(2.8%)
Spain	Yes	No (1.1%)	No (1.5%)	(3.0%)
France	Yes	Yes	Yes	(0.3%)
Italy	No (0.6%)	Yes	No (0.2%)	(0.8%)
Cyprus	Yes	Yes	Yes	(1.0%)
Latvia	No (1.4%)	Yes	Yes	(1.3%)
Luxembourg	Yes	No (0.2%)	No (0.3%)	(0.9%)
Hungary	Yes	Yes	Yes	(1.2%)
Malta	Yes	Yes	Yes	(0.3%)
Netherlands	Yes	Yes	No (1.2%)	(1.4%)
Austria	Yes	No (1.0%)	No (1.4%)	(3.0%)
Poland	No (1.4%)	No (2.3%)	No (2.4%)	(3.2%)
Portugal	Yes	Yes	Yes	(0.2%)
Slovenia	Yes	Yes	Yes	(1.1%)
Slovakia	Yes	Yes	Yes	(0.3%)
Finland	Yes	Yes	Yes	(0.2%)

Notes: In parentheses, the coefficient of variation of the weighted total. For gender and age, the average coefficient of variation over the categories is shown. Age groups are: less than 25, 26 to 44, 45 to 64, 65 and over.
*For age, only the coefficient of variation on the standard age categories is shown, since different age groupings were used in different countries to calibrate replicate weights.

7.2.5 Extension to multi-stage sampling

In each stage, the sampling of units (primary, secondary, and so on, up to ultimate) induces an additional component of variability. In multi-stage designs, the usual assumption in this case is that the sampling variance comes mostly from the first stage of sampling (i.e. the selection of PSUs and not the selection of secondary sampling units (SSUs) in each PSU). This allows both a simplification of variance formulae and a reduction of the computation burden (although this does not apply to the bootstrap), with a negligible loss of information in the presence of small sampling fractions in the subsequent stages.

The approach proposed by Preston (2009) is an alternative. This is an extension of the without-replacement bootstrap to multistage sample designs. Osiewicz and Pérez-Duarte (2012) apply the same methodology in the case of a with-replacement bootstrap, making it a direct extension to the Rao-Wu bootstrap. It is applicable to multi-stage stratified sample designs where the sampling fraction at the first stage is not negligible. Its use is transparent to final users of the data, since all the information is included through the replicate weights. The multi-stage rescaled bootstrap shows an improved estimation of the variance when two stages are used in the calculation of the replicate weights, but the gain of a third stage is minor.

7.3 Combining replicate weights and multiple imputation

In the description below, we consider the general features of a multiply-imputed sample survey, as is described in Chapter 6 of this report. Each observation has a final estimation weight w_i. There are M implicates (multiple imputation) indexed by m, and B replicate weights w_{ib} indexed by b. In the HFCS, $M = 5$ and $B = 1000$.

For each implicate m, the estimator of interest θ_m is calculated using the estimation weight w_i (for example the population total of a variable y, as $\sum_i w_i y_{im}$). The variance of this estimator is estimated using the bootstrap weights as follows: for each of the B replicates, using the replicate weight w_{ib}, calculate θ^*_{mb}, with mean across replicates $\bar{\theta}^*_m = \frac{1}{B}\sum_{b=1}^{B}\theta^*_{mb}$. The partial variance for implicate m is $U_m = \frac{1}{B-1}\sum_{b=1}^{B}(\theta^*_{mb} - \bar{\theta}^*_m)^2$. This is the standard bootstrap variance used in complete case analysis.

The total variance is then calculated according to the MI formula

$$T = W + \left(1 + \frac{1}{M}\right) Q,$$

where W is the within variance $W = \frac{1}{M}\sum_{m=1}^{M} U_m$ and Q is the between-imputation variance, $Q = \frac{1}{M-1}\sum_{m=1}^{M}(\theta_m - \bar{\theta})^2$ and the final estimator of interest is $\bar{\theta} = \frac{1}{M}\sum_{m=1}^{M}\theta_m$.

7.3.1 Test statistics

According to multiple imputation theory, the quantity $(\theta - \bar{\theta})T^{-\frac{1}{2}}$ is approximately distributed as a t-distribution with ν_M degrees of freedom, with $\nu_M = (M - 1)\left(1 + \frac{W}{\left(1+\frac{1}{M}\right)Q}\right)^2$. Barnard and Rubin (1999) recommend an alternative measure in the case of small samples, since in that case, the ν_M can be much larger than the complete data degrees of freedom. This recommended measure is $\nu^*_M = \left(\frac{1}{\nu_M} + \frac{1}{\nu_{obs}}\right)^{-1}$, where $\nu_{obs} = \frac{\nu_0+1}{\nu_0+3}\nu_0(1-\gamma)$, ν_0 is the complete-data degrees of freedom, and $\gamma = \frac{\left(1+\frac{1}{M}\right)Q}{T}$.

In the context of sample surveys, the degrees of freedom are customarily calculated as $n - L$, where n is the number of PSUs and L is the number of strata. For the HFCS, at the euro area level as a whole, it is likely that the large sample assumption holds, and that the measure ν_M is more appropriate. However, when looking at country-level data, when the number of PSUs is not large, it may be more appropriate to use the small sample formulas. It is proposed to leave this decision to final users. The information on the number of degrees of freedom by country has been included in the HFCS metadata documentation.

7.4 Variance estimation of changes between waves

In addition to estimating variances of indicators at a given time t, the second wave of the HFCS adds the time series dimension to the data analysis. It is therefore necessary to understand the principles of estimating the variance of changes between time t and $t+1$ for different estimators. The estimator for a parameter Y at a given time t for a probability sample s_t is denoted as $\hat{Y}_t$. $\hat{Y}_t$ appropriately reflects the sampling design used to select s_t. Correspondingly, $\hat{Y}_{t+1}$ denotes the estimator for the same parameter at time $t+1$, which again appropriately reflects the sampling design used to select s_{t+1}.

The change in the estimator of parameter $\hat{Y}$ between t and $t+1$ can be denoted as $\widehat{D} = \hat{Y}_{t+1} - \hat{Y}_t$. The variance of $\widehat{D}$ is given by:

$$\mathrm{Var}(\widehat{D}) = \mathrm{Var}(\hat{Y}_t) + \mathrm{Var}(\hat{Y}_{t+1}) - 2\mathrm{Cov}(\hat{Y}_t, \hat{Y}_{t+1}),$$

where $\mathrm{Var}(\hat{Y}_t)$ and $\mathrm{Var}(\hat{Y}_{t+1})$ denote the unconditional variances of $\hat{Y}_t$ and $\hat{Y}_{t+1}$ respectively, and $\mathrm{Cov}(\hat{Y}_t, \hat{Y}_{t+1})$ denotes the unconditional covariance between $\hat{Y}_t$ and $\hat{Y}_{t+1}$.[32] When the sampling designs at time t and $t+1$ are statistically independent, the estimators of the parameter Y are also independent. Consequently, the covariance between the two estimators of parameter Y is 0 and the variance of the change in the parameter is equal to the sum of variances of $\hat{Y}_t$ and $\hat{Y}_{t+1}$. If the two samples are not statistically independent, usually $\mathrm{Cov}(\hat{Y}_t, \hat{Y}_{t+1}) > 0$ and the estimates of change are more efficient.

The HFCS includes samples that have a panel component, which means that the cross-sectional samples of t and $t+1$ are not statistically independent. On the other hand, there are no instances where the net samples at t and $t+1$ would consist of exactly the same population, due to refresher samples, attrition and other types of entries to and exits from the sample population.

While it is important to acknowledge the impact of sample coordination on the variance of changes in parameter values, calculating exact measures of such variance is far from being trivial. There is no universally recognised methodology for the estimation of the covariance between $\hat{Y}_t$ and $\hat{Y}_{t+1}$.[33]. Furthermore, taking the covariance between these estimators as zero in two household surveys conducted with identical sampling designs at different times will lead to conservative estimates of the precision of changes and overstate variance.

7.5 Software routines for estimating total variance

Most good quality statistical software packages include routines for using multiply imputed data, and most also include routines for datasets with replicate weights.

---

[32] See Eurostat (2013).

[33] Several papers (see e.g. Berger, 2004; Berger and Priam, 2010) propose methodologies to estimate covariance matrices for estimators measured at different points of time for overlapping samples using various kinds of information on sampling designs.

However, not many have directly usable routines for taking into account both components of total variance. In this section, we describe a number of routines in Stata and SAS.

7.5.1 Application in Stata

Stata has had an official system for dealing with multiply imputed data since version 11, called mi. It also has procedures for using bootstrap replicate weights using the standard svy command, starting with version 11.1. From version 12 on, there is an undocumented procedure for combining both elements of the variance estimation. The mi command has a mi svyset command, which accepts replicate weights, but the mi estimate: svy: command does not allow bootstrap weights unless used with the option “vceok”.

Table 7.2
Stata code for the HFCS multiply imputed dataset

```
/* import the data */
mi import flong, m(im0100) id(sa0100 sa0010)
/* set the survey weights and bootstrap weights */
mi svyset [pw=hw0010], bsrweight(wr0001-wr1000) vce(bootstrap)
/* estimation of mean and variance */
mi estimate, vceok esampvaryok: svy: mean da1110
```

7.5.2 Application in SAS

The SAS statistical system has several routines starting with version 9.1, which allow the estimation of variance under multiple imputation and replicate weights. The core routines are PROC SURVEYMEANS (and the related ones in the SURVEY… family of procedures) and PROC MIANALYZE.

The example below shows how the mean of the derived variable DA1110 can be calculated, and how a linear regression could be run.

Table 7.3
SAS code for the HFCS multiply imputed dataset

Means

```
proc surveymeans data=HFCS varmethod=brr(fay=0.000);
     var da1110;                  * variable of interest;
      repweights wr0001-wr1000; * replicate weights;
         by im0100;                  * implicates;
      weight hw0010;              * estimation weight;
             ods output Statistics = outex1 ;
                        run;

            proc mianalyze data=outex1;
                   modeleffects mean;
                     stderr stderr;
                        run;
```

Regression

PROC MIANALYZE expects the input dataset to contain either one line per implicate, or a variable called _Imputation_. The IM0100 of the HFCS thus needs to be renamed.

```
proc surveyreg data=HFCS varmethod=brr(fay=0.000);
          model da1110 = da1120;    * model;
     repweights wr0001-wr1000; * replicate weights;
        by im0100;                 * implicates;
     weight hw0010;            * estimation weight;
        ods output ParameterEstimates = outex2 ;
                        run;

          proc mianalyze parms=outex2;
           modeleffects intercept da1120 ;
                        run;
```

8 Statistical disclosure control

Statistical disclosure control for the HFCS has two facets: safe data and safe users. The latter refers to the procedure for granting access to the HFCS dataset, such as the confidentiality declaration necessary before the data can be disseminated to third parties. The former is the process by which the data collected during the survey are anonymised, i.e. are treated in such a way that the effort necessary to re-identify a particular respondent, either a household or a person, is disproportionately high. This chapter deals with this anonymisation process.

8.1 General principles in the HFCS

The anonymisation procedure is applied either by the NCB (or NSI, i.e. before submitting the data to the ECB) or at the ECB level, and is designed to ensure, insofar as possible, data comparability. Country-specific anonymisation techniques may also be applied centrally by the ECB in close coordination with the NCB (NSI) concerned, to ensure the confidentiality of responses where necessary.

The anonymisation procedure has two main components: a “general procedure” and “country-specific modules”. The general procedure is applied to the data of all countries, while country-specific modules, imposed by different data protection regulations, different assessments of disclosure risk or different traditions, are applied on a case-by-case basis, where needed.

In addition, more information than provided for in the general procedure may be included in the dataset. In that case, as many variables as required containing the additional information are added to the research dataset.[34]

It consists of the following techniques:

- The following variables are kept unchanged: country and type of dwelling. In the case of a panel survey, the following variables are kept unchanged: vintage of last interview and survey vintage. In addition, unique household identification numbers in a randomised form for the current and past (in the case of a panel) survey wave are kept unchanged. If they are not provided in a randomised form by the Member State, the ECB will randomise them before dissemination. The last interviewer’s call date is recorded by the quarter in which it took place. All other variables relative to the sample are deleted.
- Only those households that participated in the survey are included in the research dataset (according to the survey database outcome variable); non-respondents are not included.

[34] For example, the file contains two versions of the variable HB0100 (size of main residence in square metres), one as a continuous measure (only for those countries where releasing such information does not pose substantial disclosure risks), the other in brackets of 10 square metres.

8.1.1 Top-coding and deletion of variables

This section only lists the major perturbations that have been applied to the collected information, as described in the documentation for the microdata (UDB documentation documents 1 to 5, available on the ECB website). The full list of changes is available in Appendix 11.5.

Demographics

Age is top-coded at 85 years. In Ireland and Malta, only age in five-year brackets is provided in a separate variable. Due to the top-coding, several other variables related to age have been either top- or bottom-coded (e.g. how long has the household been living in their main residence).

Country of birth is recoded in four categories, showing only the country where the survey took place, other euro area countries, other European Union countries, and other countries. This also applies to the non-core variable Country of citizenship.

Education is coded in four categories, according to the International Standard Classification of Education (ISCED), version 1997, namely ISCED1, ISCED 2, ISCED 3+4 and ISCED 5+6. This also applies to the non-core variable Education of the parents.

Real assets

In addition to age-related coarsening, the size of the household main residence is bracketed into ten categories in three countries. The number of employees in self-employment businesses owned by the household is bracketed into four categories in several countries.

Employment, Pensions & Inheritances

Only age-related coarsening has been applied.

8.1.2 Additional bracketing

In addition to the changes to the variables described above, in some countries, a number of additional variables have been top-coded or recoded into coarser categories in order to reduce identity disclosure risk.

8.2 Collapsing of cases

In the case of very rare assets, different variables might be collapsed. This is the case of boats and planes, which are grouped into the residual category in a few countries.

8.3 Random rounding

This approach is proposed in Kennickell and Lane (2007) for the US Survey of Consumer Finances (SCF).

The idea is to avoid identification through matching with amounts provided with full detail by the household. The solution is to round the numbers to a specified precision, randomly, in a way that does not bias the results (either up or down, based on how far the amount is from the rounded values above and below).

This procedure is equivalent to adding random noise of mean 0 to each amount, with heteroscedastic variance. For example, 12,345 would get rounded to 12,000 approximately two-thirds of the time, and to 13,000 one-third (if we are rounding to two digits). This is done independently across implicates.

Altogether, this is a minor measure of statistical disclosure control whose effect is limited, as respondents often spontaneously round many amounts. It only needs to be applied when there is a clear case of re-identification risk (e.g. matching with administrative data). Internal tests have shown that rounding to two digits has a minimal effect on sample means, while, when rounding to three digits, the effect is also minimal on medians.

Random rounding to three digits was applied to certain variables in Estonia, namely the amounts outstanding of credit lines and overdrafts, and values of sight and savings accounts, mutual funds, bonds, publicly traded shares, social security plans and voluntary pension plans, and income from public pensions, unemployment benefits and social transfers.

Table 8.1

Rounding of variables in nominal amounts

Data range (USD in the SCF, EUR in the HFCS)	SCF rounding to the nearest…	Rounding to two digits, to the nearest…	Rounding to three digits, to the nearest…
>1 million	10,000	100,000	10,000
100,000 to 1 million	1,000	10,000	1,000
10,000 to 100,000	1,000	1,000	100
1,000 to 10,000	100	100	10
100 to 1,000	10	10	1
5 to 100	10	1	1
-4 to 4	1	1	1
-5 to -100	10	1	1
-100 to -1,000	10	10	1
-1,000 to -10,000	100	100	10
-10,000 to -100,000	1,000	1,000	100
-100,000 to -1 million	1,000	10,000	1,000

Source for the SCF column: rounding used for most of the variables in the 2010 wave of the SCF. Data bottom-coded at -1 million. Some variables (e.g. hourly wages) receive a slightly different rounding treatment and are not reported here.

9 Comparability issues

One of the goals of the HFCS project is to ensure as much as possible that the data will form a homogeneous set. While much effort was spent in trying to achieve this consistency, such an ambitious exercise covering diverse countries, markets, structures and cultures will probably suffer from some comparability issues. This may make it difficult to disentangle the extent to which cross-country variation is due to such structural divergences as opposed to other economic, financial and/or psychological factors influencing household decisions.

This chapter does not attempt to draw an exhaustive list of all such issues, but just to highlight the most relevant ones with a view to helping users better understand what is behind the data.[35]

9.1 What are comparability issues?

When analysing data coming from the HFCS, users want to know to what extent they can draw conclusions from cross-country differences, in other words, to what extent apparent differences are real rather than an artefact of measurement.

9.2 Dimensions in the assessment of comparability

Comparability issues could be classified in various sets. Differences between countries can result from timing, survey mode, questionnaire, editing, imputation and anonymisation.

9.2.1 Time dimension

There are several dimensions in the treatment of the time comparability of the survey. The most immediate one is the fieldwork period, i.e. when and for how long the data were collected in each country. The length of the fieldwork is indeed important, as economic conditions may have significantly changed between the beginning and the end of the fieldwork period. Finally, another important factor which may trigger comparability issues is the reference period for wealth (assets and liabilities, as stocks at a particular point in time) as well as income (flow of income over a period of 12 months).

[35] The status of each variable in each observation of the HFCS is coded in a “flag variable”, available in the microdata. It codifies whether the variable is missing (and why), was recorded as provided in the data, or has been edited, imputed or estimated. Flag variables are thus an extremely rich source of information at the granular level on data issues, and users are urged to take this information into account when analysing the data.

All these components play a role in the comparability of the data, and should be kept in mind when comparing different country results.

The fieldwork in most countries ranges from March 2013 to March 2015. The reference periods for assets and liabilities range in most cases from the first quarter of 2013 to the beginning of 2015. The reference periods for income cover 2012, 2013 and 2014, or the 12 months before the interview.[36]

Table 9.1

Reference periods and inflation adjustment factor between the 1st and 2nd wave

Country	Fieldwork	Assets & Liabilities	Income	Inflation adjustment factor between 1st and 2nd wave
Belgium	06/2014 - 01/2015	Time of interview	2013	1.079
Germany	04/2014 - 11/2014	Time of interview	2013	1.072
Estonia	03/2013 - 06/2013	Time of interview*	2012	-
Ireland	03/2013 - 09/2013	Time of interview	Last 12 months	-
Greece	06/2014 - 10/2014	Time of interview	Last 12 months	1.066
Spain	10/2011 - 04/2012	Time of interview	2010	1.049
France	10/2014 - 02/2015	Time of interview	2014	1.081
Italy	01/2015 - 06/2015	31/12/2014	2014	1.079
Cyprus	02/2014 - 07/2014	Time of interview	Last 12 months	1.068
Latvia	04/2014 - 09/2014	Time of interview	2013	-
Luxembourg	04/2014 - 12/2014	Time of interview	2013	1.093
Hungary	10/2014 - 11/2014	30/09/2014	1/10/2013 - 30/09/2014	-
Malta	01/2014 - 06/2014	31/12/2013	2013	1.069
Netherlands	04/2014 - 03/2015	31/12/2013	2013	1.091
Austria	06/2014 - 02/2015	Time of interview	2013	1.101
Poland	01/2014 - 02/2014	Time of interview	2013	-
Portugal	03/2013 - 07/2013	Time of interview	2012	1.079
Slovenia	09/2014 - 12/2014	Time of interview	2013	1.074
Slovakia	02/2014 - 04/2014	Time of interview	2013	1.094
Finland	01/2014 - 05/2014	31/12/2013	2013	1.108

Source: HFCS metadata.
*Time of interview for variables collected in the interview, 30/4/2013 for variables derived from register data.

The time dimension has an effect on comparability, since the amounts shown for the second survey wave are nominal and do not include any adjustments for inflation. However, the figures between the two survey waves in individual countries have been adjusted for inflation using the Harmonised Index of Consumer Prices. The values of assets, debt, income and consumption have been adjusted for by multiplying the first-wave figures with the ratio between the yearly averages of the price level between the reference years in the two waves of the survey. The adjustment factors are shown the right hand column of Table 9.1. An adjustment factor of 1.079 indicates that inflation between the two survey waves was 7.9%.

[36] In Spain, the fieldwork period was between October 2011 and April 2012, the reference period for assets the end of 2011 and the reference period for income the year 2010.

Overall, inflation is only one source of variability over time. Housing and financial market developments over the course of the fieldwork period have impacted the value of household assets, and have altered the comparability of figures not only across countries, but also within countries over the duration of the fieldwork, in particular in cases of rapid price movements. It was decided not to correct the amounts reported in the report on the results of the second wave for inflation, as such a correction would, first of all, not change any of the conclusions, and second, introducing this correction may give readers the incorrect impression that owing to the adjustment, the data are more comparable than they are in reality.

9.2.2 Purchasing power parity

A much bigger difference concerns the differences in "cost of living" across countries, usually expressed in purchasing power parities (PPP). These corrections are meaningful when concerning consumption-related values or living standards (for example, income). However, the rationale for adjusting wealth figures using PPP is not clear, and has not been used in reporting the results of the survey. Table 9.2 shows the purchasing power parities in various countries at the date of reference, which in this case indicates the time of interview or the reference period for balance sheet items, where applicable.

Table 9.2

Possible Inflation and purchasing power parity correction factors, second wave

Country	Date of reference	PPP
Belgium	2014	0.802
Germany	2014	0.859
Estonia	2013	1.154
Ireland	2013	0.713
Greece	2014	1.021
Spain	2011	0.945
France	2014	0.810
Italy	2014	0.848
Cyprus	2014	0.968
Latvia	2014	1.244
Luxembourg	2014	0.724
Hungary	2014	1.517
Malta	2013	1.078
Netherlands	2013	0.794
Austria	2014	0.824
Poland	2014	1.563
Portugal	2013	1.066
Slovenia	2014	1.068
Slovakia	2014	1.286
Finland	2013	0.713

Notes: HICP: Harmonised Index of Consumer Prices, Overall index, calculated to adjust values in the HFCS to 2014 amounts.
Source: Eurostat (2015) for HICP and purchasing power parity factors, HFCN calculations.
How to read: euro amounts in Belgium should be multiplied by 0.802 and by 1.154 in Estonia to correct for PPP differences.

9.2.3 Sampling and survey mode

As seen in Chapter 3, sampling in most countries is carried out by personal, face-to-face interviews, with the aid of a computer (CAPI). In three countries, collection is by other means (telephone, paper and pencil interview and web), with a small fraction of interviews in two other countries with face-to-face paper questionnaires. Finally, in one country, there is a predominance of paper questionnaires.

9.2.4 Questionnaire

The questionnaire was translated and adapted into the local language(s) by each institution. In many cases, adapting one question of the common questionnaire required several questions in the local questionnaire to capture the different facets of the issue in the local culture. Although special care was taken to ensure the accuracy of this step, this adaptation process may have led in some cases to slight differences in the output result.

The common questionnaire is already completely implemented in various countries, and the share of collected variables has increased compared with the first HFCS wave. A total of 451 variables in the household-level file and 61 in the person-level file were envisaged. In the household-level file, these variables refer to 157 different items with various numbers of loops and secondary purposes,[37] and in the person-level file to 52 different items. Since various loops and e.g. secondary purposes of loans were applicable only to a very limited number of households or persons, the table below shows the items rather than the output variables collected in various countries.

Most countries that provide the least number of core variables have adapted the HFCS to an existing survey and the process towards full harmonisation with the HFCS is still ongoing. Consequently, these countries also provide more non-core variables than others (see Appendix).

[37] An item indicates, for example, the outstanding value of one mortgage on the household main residence or the purpose of this mortgage. Many items are collected in loops e.g. for the three most important loans (see section 2.4.1), and up to nine different purposes can be given by the household.

Table 9.3

Items available in the User Database (UDB)

Country	Household-level file (maximum 157)	Personal-level file (maximum 52)
Belgium	157	52
Germany	156	51
Estonia	152	49
Ireland	154	52
Greece	157	44
Spain	141	43
France	147	47
Italy	141	47
Cyprus	157	52
Latvia	157	49
Luxembourg	157	52
Hungary	156	44
Malta	157	50
Netherlands	154	49
Austria	157	45
Poland	138	42
Portugal	156	52
Slovenia	157	49
Slovakia	157	51
Finland	105	35

Notes: the table displays the number of variables with at least one non-empty observation.

Due to questionnaire differences, some variables cannot be provided with the same amount of detail in the microdata. This is the case for occupation (according to the ISCO-08 classification, provided on one or two digits) and activity (according to the NACE classification, at the section level, with some sections grouped in some countries).

9.2.5 Income

The core output variables on income are defined in gross terms. However, there were different approaches for the collection of income. In ten countries, income was collected in gross terms only. In Italy, net income was collected and gross income constructed by estimating the amount of taxes and social contributions with the help of legislative and institutional parameters. Respondents had the option to provide net income for all income components in Germany, Greece, Poland and Portugal, and for some income components in Belgium, Latvia and Austria (see Table 9.4), in which case gross income was estimated. Estonia and Finland had access to income registers and provided taxes and social contributions in addition to gross income, which enables the calculation of net disposable income[38]. In France all income components are based on information from tax income registers and in Ireland

[38] The concept of "net income" varies country by country, and has not been harmonised.

register data was used in addition to interview data to derive several income variables.

Information on which observations were estimated in this way is recorded in the flag variables for each income component, using the flag value 5050 (Estimated, originally not collected).

Table 9.4

Deviations in the collection of income variables

Country	Information
Belgium, Latvia	If respondents were not able to provide gross amounts, net income was collected for employee, self-employment and pension income, and gross amounts estimated.
Germany, Portugal	Gross income collected, but respondents had the option to provide net income figures. If provided, net income figures were converted to gross income using information from the tax system.
Ireland	Register data on income, including employee, profits and social transfers such as unemployment benefits and pensions etc., was used in the derivation of income.
France, Finland	Income data derived from administrative sources.
Estonia	Income data from public transfers, unemployment benefits, and Estonian public pensions were derived from registers. If provided, net income figures were converted to gross income using information about the tax system.
Spain	Information on public pensions also includes private pensions; information on private pensions was not collected separately.
Greece	The respondent was able to choose whether to provide the income figure gross or net and its frequency (monthly or annual). The net amounts were converted to gross by adding the estimated social contributions and the estimated applicable tax. The applicable tax rate on income is calculated on the basis of 2013 tax regulation.
Italy	Income is always collected net of taxes and social contributions. Gross incomes are reconstructed using a methodology developed on the basis of information on personal income tax and social contributions at national level and on the basis of the demographic characteristics of the household members. This methodology is different from that used in wave 1. The gross amounts should not be compared between the two waves, only net incomes are comparable. Income from financial investments not directly collected, but calculated using average interest rates and information collected on households financial assets
Austria	If respondents were not able to provide gross amounts, net income was collected for employee, self-employment and pension income as well as income from financial assets. These net income data were transformed to gross income using information of net income, employment status, household structure and geographical location of the household in combination with tax system information.
Poland	If respondents were not able to provide gross income amounts, net income was collected. Gross income was then imputed, using net income as one of the covariates.

Source: ECB – HFCS metadata.

9.2.6 Editing

As described in Chapter 6, the purpose of editing is to manually correct cases where the information has been erroneously recorded.

The major reason for editing in the HFCS is the conversion from net to gross income in countries where the information was collected in net terms. This conversion takes the form of a model, specific to each country, date, employment status and household structure. In the absence of sufficiently detailed information, which would be prohibitively expensive to collect in a face-to-face survey, the conversion requires a number of assumptions, which might limit comparability not only across countries, but also across households within each country.

Financial income, i.e. income earned from financial assets, was estimated in Italy, given that attempts to collect this information directly from households often meet with little success.

9.2.7 Imputation

In order to calculate reliable country- and euro area-level information, the HFCN defined a set of variables that were to be imputed by all participating institutions (including variables on possession and values of assets, liabilities, and income). Nevertheless, due to a combination of factors, this was not always possible.[39] Table 9.5 lists the number of variables in the HFCS core variables in the to-be-imputed list that contain more than five households or persons whose value should have been imputed but was not, or that were collected only for a part of the sample. This table does not include variables that were not collected at all in individual countries, and which are shown in Table 9.3.

Table 9.5

Number of variables in the to-be-imputed list with missing values

Country	Household-level file	Personal-level file
Belgium	2	0
Germany	0	1
Estonia	0	0
Ireland	2	1
Greece	0	0
Spain	0	0
France	65	6
Italy	0	1
Cyprus	0	0
Latvia	0	0
Luxembourg	0	0
Hungary	6	0
Malta	0	0
Netherlands	7	0
Austria	0	0
Poland	21	4
Portugal	0	0
Slovenia	0	0
Slovakia	0	0
Finland	1	0

Source: ECB – HFCS metadata.

[39] See Chapter 6 for further details. In some cases, differences in the national implementation of the HFCS questionnaire lead to cases that cannot be imputed: see Section 9.2.4.

9.2.8 Anonymisation

As discussed in Chapter 8, although a core set of common anonymisation procedures has been applied to all country surveys, in order to protect the anonymity of respondents, and in agreement with national practices, additional steps have been applied in some countries. Care has been taken to provide researchers with a set of less common variables, for example, in the case of age (coarsened to five-year brackets in some countries), by providing the coarsened variable for all countries.

10 Comparability between the HFCS and other statistics

The HFCS provides a unique data source on household-level wealth, indebtedness, income and consumption, for the euro area, Hungary, and Poland. While this kind of data, where all these topics are covered by one data source at the individual level, are not available elsewhere, individual components of the survey are measured by other statistics. The definitions of variables and data production approaches are sometimes, though, quite different compared with those used in the HFCS.

This chapter shows comparisons between the results of the HFCS and other statistics. First, the demographic structure is compared with other data sources producing personal- or household-level information. Subsequently, the core of the HFCS, data on wealth and liabilities, is compared with macro sources. Finally, the comparability of the income data is assessed.

10.1 Comparability of the demographic structure of the HFCS

The target population of the survey are private households residing in the national territory at the time data are collected and their current members. For the results of the survey to be reliable, it is essential that the structure of the survey population by age, household size, economic activity, etc. is coherent with the target population.

In a sample survey, the structure of the population is determined by sampling and weighting procedures, described earlier in this document. In the sampling stage, it is crucial that a sufficient number of members of relevant population groups are included in the sample of households that are interviewed. After the data have been collected, sample weights are constructed, and as a result each household providing data to the survey is designated to represent a certain number of households in the target population. Consequently, the sum of weights of units belonging to selected population groups indicates the size of these groups in the survey population.

A variety of external sources measure the structure of the household population in each euro area country. The first benchmark source used in this report is population statistics by Eurostat, which is available in each EU country for the survey reference periods. Population statistics provide accurate measures of the population size, along with several breakdowns, e.g. by age and gender.

Population statistics enable the comparison of basic personal-level data. For comparison of household-level data with identical definitions of households, as well as for some more detailed individual level characteristics, data from other surveys are the only feasible benchmark. In this chapter, HFCS data are compared with EU Statistics on Income and Living Conditions (EU-SILC), which is a harmonised survey conducted annually in every EU country. When comparing the two surveys, it should be kept in mind that EU-SILC faces the same challenges of a household survey like

the HFCS, and differences between the outcomes of these two data can be caused by methodological issues in either of the two surveys.

In the following chapters, the demographic structure of the HFCS data is compared with external benchmarks with respect to age, household size and labour status.

10.1.1 The role of calibration and the comparability of demographic statistics

The demographic structures produced at the country level are greatly affected by the selection of variables and sources used in the calibration of weights (see Table 5.2). Age is used as a calibration variable in 18 countries, household size in 16 and labour status in 6 countries.

In a few cases, both the variables and source data used in calibration exactly match the EU-SILC basic demographic statistics. This is the case for age in Belgium and Portugal and for household size in Estonia and Greece (for three categories). The Finnish data are based on the EU-SILC sample, and only the addition of a number of variables explaining the distribution of wealth (register data on mutual funds and listed shares) to the calibration of HFCS weights causes minor differences with the demographic structure of EU-SILC.

Harmonised population and housing census data was produced for 2011 in all EU countries. The census is frequently used as the basis of benchmark statistics and the data produce very similar demographic structures. Several countries use census data for calibration. While census data provide more detailed information on the household structure, the data are only produced at ten-year intervals. Consequently, the reference period of the data used for calibration may be different to that of the benchmark data. For example, in Spain, the 2011 census data became available only after the publication of the HFCS results.

In some countries, the EU-SILC statistics have different definitions of the target population than the country-specific HFCS. In Austrian population statistics and EU-SILC, the definition of household population only includes households that live in dwellings, in which a main residence is officially registered. The Austrian HFCS sample includes households that live in dwellings where no main residence is officially registered. Consequently, these statistics are not used as a benchmark for weight calibration, and the difference of household definition should be taken into account when comparing these two sources.

10.1.2 Age structure

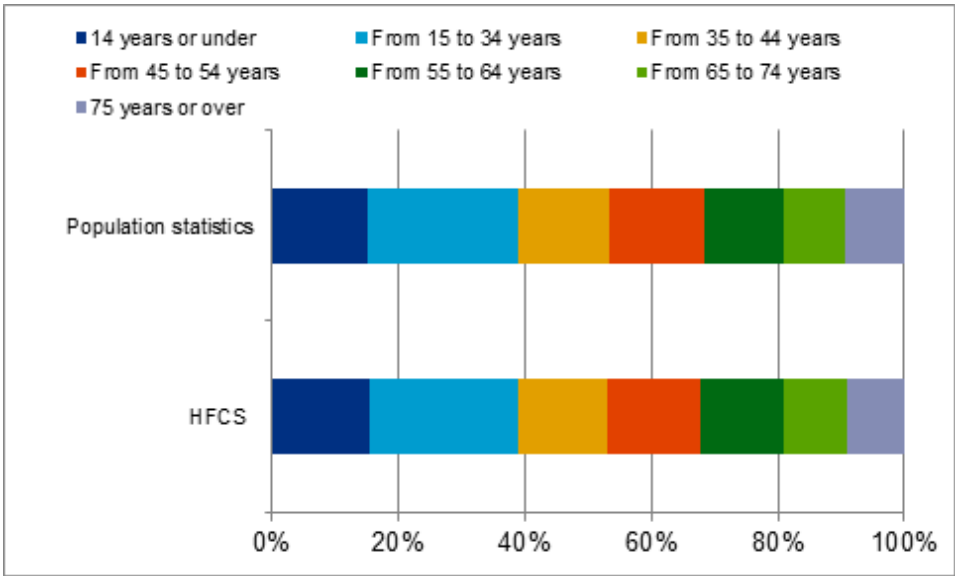
The development of net wealth follows a hump-shape profile over the age of the household reference persons. Net wealth rises approximately to the age of 60, and declines gradually thereafter. Wealth differences between the youngest age groups and the age groups close to retirement age are substantial. It is therefore crucial that

the survey population by age provides a good representation of the target population.

Figure 10.1 shows the age structure of persons in the HFCS and population statistics. Note that this age structure is different from that used in the reporting of the results, where wealth data are analysed at the household level and the age structure shown in the results is determined by the age of the household reference person. Table 10.1 shows the age structure of all household members, including children. The age structure of the total adult population is on average younger, because younger household members are less frequently classified as reference persons e.g. in households that comprise several generations.

Figure 10.1

Euro area population structure by age in the HFCS and population statistics



Sources: ECB – HFCS and Eurostat – Population statistics.

The age structure of persons in the survey population is a very close match to the corresponding structure of population statistics in the euro area. In the HFCS, there is a slight underrepresentation of young working-age adults, while the share of the oldest working-age group between 55 and 64 years old is 0.5 percentage points higher than in the population statistics. Overall, the differences in the euro area age structures between the two statistics are small, and should not cause any significant bias in the interpretation of the results.

10.1.3 Household size

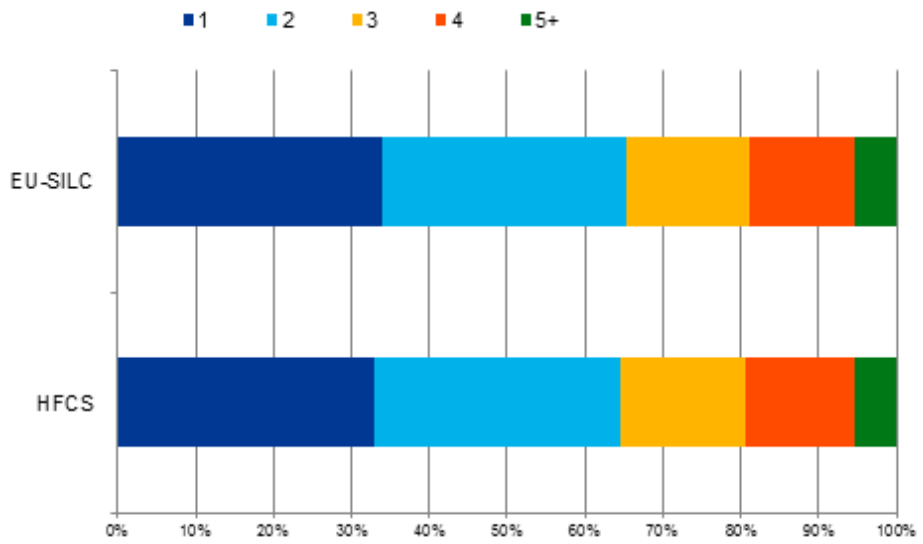
Wealth in the HFCS is reported at the household level and no equivalence scales are used, as in most income distribution statistics, such as EU-SILC. This is consistent with international recommendations on having households as the preferred unit of analysis for household wealth statistics (OECD, 2013).Therefore the distribution of the survey population by household size is an important aspect, not

only in the comparison of wealth levels, but also in assessing the representativity of the sample. Bigger households hold on average more wealth than smaller households. This is obviously driven by the larger number of adult members with wealth holdings. Additionally, larger households tend to live in larger and more valuable homes. This is crucial to acknowledge, given the significance of real assets, particularly of the household main residence, in the wealth portfolios of households.

While the definition of age is straightforward in any statistics, the definition of household is different in survey data compared with statistics based on administrative data or census data, in which the household-dwelling concept is applied (Eurostat, 2011b). In the HFCS, persons living in the same dwelling can belong to one or more different households, or one household can consist of individuals registered in different dwellings. The household composition, as defined in the HFCS, can only be determined during the interview. Consequently, it is feasible to compare the household size distribution using another survey statistics with identical household definition as a benchmark.

Figure 10.2

Euro area household structure by household size in the HFCS and EU-SILC



Sources: ECB – HFCS and Eurostat – EU-SILC.

Compared with EU-SILC, the HFCS produces a smaller share of single-person households and a slightly higher share of households with two-four members in the euro area (see Figure 10.2). The difference in the share of one person households is one percentage point. All countries, except France, Italy, Latvia and the Netherlands included household size as one of their calibration variables. Of these countries, the difference to the EU-SILC statistics is significant only in Italy. The population registers in Italy used by EU-SILC are not considered to reflect the current household size composition, e.g. due to the failure to record legal immigrants moving out of Italy during the crisis. In Slovenia, there is a notable difference between the HFCS and EU-SILC, but the HFCS household size distribution is coherent with their national statistics on households, used as a benchmark in calibration.

10.1.4 Labour status

Another important determinant of household wealth is labour status. The HFCS collects information about the labour status of each household member aged 16 or over. This variable indicates whether the person is working, retired, unemployed, and so on. For persons working, there is an additional question on whether the person is an employee or self-employed.

According to the HFCS results, households with a self-employed reference person have on average the highest wealth holdings, while working age persons who are not economically active have the lowest wealth holdings. The labour status structure has, thus, significant implications for the results.

As in the case of household size, the only comparable benchmark statistics on labour status distribution are other surveys. In EU-SILC, information on self-defined current economic status is collected with one question, with classification similar to that in the HFCS. The only differences are: self-employed and employees are defined as different categories in one question, and the category "on maternity/sick leave" does not exist as such. Persons belonging to the latter category are in most cases classified as employees in EU-SILC.

Figure 10.3 shows the distribution of the survey population aged 16 or over in the HFCS and EU-SILC by self-defined labour status. As in the case of age, this classification is done at the person level, not by the household reference person. The breakdown by labour status in the HFCS results report is based on the labour status of the household reference person, and is thus different from the breakdown presented here.

The population structure by labour status in the euro area is extremely coherent with the benchmark statistics, especially if one assumes that most of the persons classified under category "on sick/maternity leave" would be classified as employees in EU-SILC. There is a slightly higher share (0.4 percentage points) of self-employed persons in the HFCS and a slightly smaller share of the group other inactive (not working and not retired).

In five countries (Germany[40], France, Luxembourg, Hungary and Slovakia), labour status was used in the calibration, with various data sources applied as benchmarks. Additionally, in Finland, both stratification and calibration use information on different income variables that can be considered proxies of labour status. Compared with other demographic structures analysed in this report, there are more differences between the labour status structures of the HFCS and the EU-SILC statistics in individual countries. This was to be expected, since self-defined labour status is a more difficult concept for households to evaluate than age or household size.

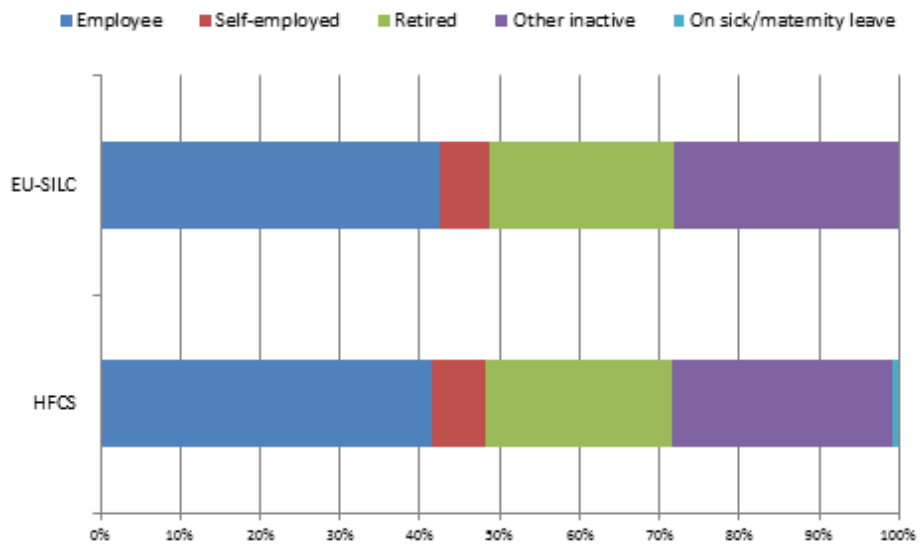
The differences in the labour status structure are caused by various methodological choices across both the HFCS and EU-SILC. Oversampling the wealthy in the HFCS

[40] In Germany, the labour status of the main income earner, not that of all persons, was used in the calibration of household weights.

is also likely to have an influence on the results. A thorough analysis of the causes for differences would require deeper research. For example, in the case of the self-employed, the fact that the HFCS collects detailed information on self-employment businesses before the question on labour status may have an impact. The role of interviewers should also be emphasised. In two cases (German EU-SILC and Dutch HFCS), data are collected by self-administered interviews. In both cases, the survey that uses interviewers produces a clearly higher share of self-employed persons.

Figure 10.3

Euro area population structure by labour status in the HFCS and EU-SILC



Sources: ECB – HFCS and Eurostat – EU-SILC.

10.2 Comparing the HFCS and macro data on financial wealth and liabilities

Data on household sector wealth and liabilities are also available in national accounts and other macro sources. While it is useful to compare wealth data from micro and macro statistics, it must be kept in mind that there are significant differences between the definitions and methodologies applied in the two statistics. Consequently, differences in the levels of wealth between the two data sources are expected to be observed, especially if one compares the concepts of aggregate wealth used in each source.

There are several reasons for the discrepancy between total wealth levels derived from micro and macro sources. Coming from different traditions and addressing different purposes, the micro and macro approaches have developed quite independently. Thus, there is significant variability in the practices in assessing the boundaries of the household sector, in the valuation of assets and reference periods and in the definition of wealth and individual wealth items.

Survey microdata aim to analyse income and wealth distributions, as well as compare income, wealth and debt across different sub-populations. Household-level data allow important insights into the economic behaviour of households that cannot be provided with macro level information. The main value added of household survey data is to answer relevant research and policy questions rather than to produce accurate statistics on the wealth aggregates. Conversely, survey data have to deal with issues such as possible reporting and sampling bias. Of these, the latter has a particularly large impact on wealth data, given the skewed distribution of wealth and the difficulty of contacting households at the very top of the distribution.

Macro data permit, *inter alia*, investigation into how different institutional sectors contribute to the national product, consumption and saving, as well as to national wealth. National accounts are constructed in a way that seeks to minimise bias in the estimates for the economy as a whole, as well as to minimise statistical discrepancies within the system. Therefore, some bias may be recorded in the household sector accounts to satisfy the balancing constraints of the whole system of accounts. In some cases, certain economic transactions for the household sector may even be derived as residual, by subtracting from the estimated total the estimates of other institutional sectors.

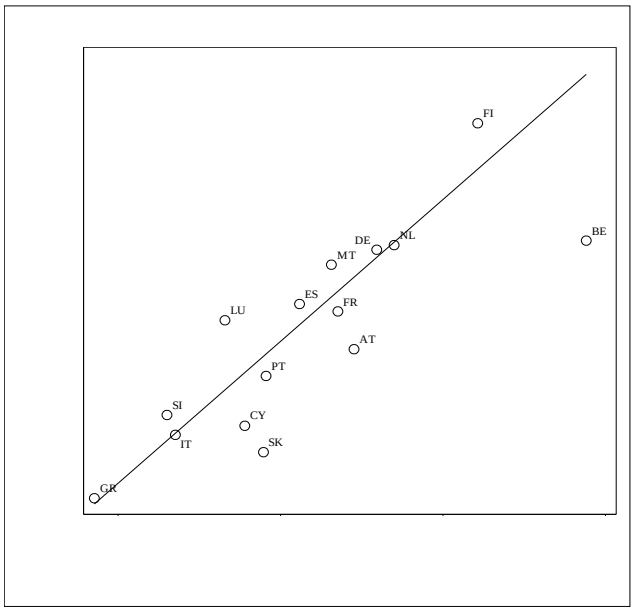
These kinds of discrepancies between micro and macro data have been analysed in recent years, e.g. by Andreasch and Lindner (2014) and Honkkila and Kavonius (2013). This document does not intend to repeat the discussions from these papers, but acknowledges the conclusions on the differences between the methodologies. Instead of analysing total (financial) wealth with the concepts applied in micro and macro statistics, this chapter concentrates on comparing wealth items that are relatively comparable across the two sources.

10.2.1 Financial assets

The data from the first HFCS wave showed that the levels of financial wealth in the survey data are generally lower than the levels produced by national accounts. The degree of underreporting of financial assets in the survey data was significantly higher than in the case of real assets, for which the HFCS in some countries produced even higher aggregate values than national accounts. There were also cross-country differences in the ratio between financial wealth in the HFCS and financial wealth in national accounts. It is fair to assume that a portion of these cross-country differences is caused by divergences in the methodologies applied in the country-level production processes of both statistics. The second-wave data allows the testing of this hypothesis by evaluating whether the ratios between the levels of wealth are stable across time in various countries.

Figure 10.4

Ratio of adjusted financial wealth per capita in the HFCS to national accounts, first HFCS wave (around 2010) and second HFCS wave (around 2014)



Sources: ECB – HFCS and ECB – Annual Sector Accounts.

According to previous literature, financial wealth items with similar definitions in surveys and national accounts are deposits, mutual fund shares, listed shares and bonds. This concept will be called *adjusted financial wealth* in the remainder of this chapter. These items are summed up for both statistics, and the ratio of HFCS totals to national accounts totals in per capita terms are shown for all countries that participated in the first HFCS wave and both survey waves in Figure 10.4.

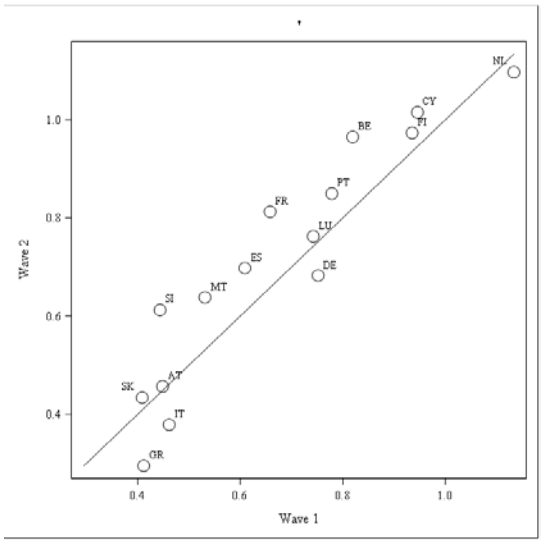
Figure 10.4 shows that the HFCS produces lower levels of per capita financial wealth than macro data, even if only comparable items are used in the comparison. There is also significant cross-country variability between the ratios of adjusted financial wealth. However, in most countries, these ratios are remarkably stable across the two HFCS waves in 2010 and 2014. This suggests that the differences between the results from the two data sources are predominantly caused by methodological and conceptual differences between macro and micro statistics. These issues may be country- and asset-specific, but do not change considerably over time. While the existence of some reporting and sampling bias in the survey data should be recognised, the impact of the bias in the results is not completely random and the survey data provides a reliable source for looking at the distributions of household wealth and their changes over time.

10.2.2 Liabilities

On the liabilities side of the households' balance sheets, there are minimal conceptual differences in the definitions between micro and macro statistics. In the HFCS, debt is collected by collateral – separately for mortgages on the household main residence, mortgages on other properties and loans not collateralised by real estate. In addition, private loans, i.e. loans from other households, are collected separately in the HFCS second wave. Households' liabilities in macro statistics are classified by the duration of the loan, and there is usually no differentiation between mortgages and other loans. However, the definition of the aggregate loans in macro statistics is almost identical to the HFCS definition. The only exception is private loans, which are usually not recorded in macro statistics. This has a limited impact on the evaluation, and the share of private loans in total household debt is approximately 1% in the HFCS second wave.

Figure 10.5

Ratio of households' liabilities per capita in the HFCS to MFI statistics, first HFCS wave (around 2010) and second HFCS wave (around 2014)



Source: ECB – HFCS and ECB – statistics on balance sheets of Monetary Financial Institutions

The benchmark data for liabilities used in this chapter is the ECB statistics on the balance sheets of monetary financial institutions (MFIs).[41] These data provide information on loans provided by monetary financial institutions, classified by the institutional sector of the lender. The statistics are harmonised at the euro area level. Data are collected directly from the institutions providing loans, and unlike the data from national accounts, are thus not subject to any balancing adjustments. MFI statistics have recently included data on loans adjusted for sales and securitisation, incorporating more comprehensive information on loans originated by MFIs but which are no longer recorded on their balance sheets. As a result of this adjustment, the data are comparable with any survey data on household debt.[42] The drawback of MFI data is that they do not differentiate between households and non-profit institutions serving households.

The results of the comparison of the levels of households' liabilities between micro and macro statistics are shown in Figure 10.5. The levels of debt produced by the survey are generally closer to the levels of macro data than the levels of adjusted financial wealth shown in the previous chapter. This is not surprising, since the sampling bias caused by having fewer of the richest households in the sample than in the population is smaller for liabilities than for financial wealth. A significant share of financial assets is held by extremely wealthy individuals, but the distribution of debt is much less skewed. However, cross-country differences in the HFCS/MFI ratio of liabilities can be observed.

As was already seen in the analysis of adjusted wealth, the difference between levels of debt in micro and macro statistics in individual countries is very stable across the two HFCS waves. This implies once again that the survey data are a reliable source for analysing distributions of households' balance sheets, and their changes over time.

10.2.3 Conclusions and the way forward

This chapter has provided only a brief comparison between micro and macro data on household wealth and liabilities. Work on understanding and quantifying the differences between the two data sources is ongoing. The ECB has launched an expert group on linking macro and micro statistics for the household sector. This

41 https://www.ecb.europa.eu/stats/money/aggregates/bsheets/html/index.en.html

42 The MFI data for Luxembourg in Figure 10.5 includes loans unadjusted for sales and securitisation, since securitised loans in Luxembourg are almost exclusively held by non-residents who are not included in the sampling frame of the HFCS.

expert group is analysing more deeply the concepts and practices of macro and micro statistics, including assets with a lesser degree of comparability such as unlisted shares and other equity and pension wealth. The expert group will publish a separate report on these linkages in 2017.

10.3 Comparison of income data between the HFCS and EU-SILC

The main purpose of the HFCS is to collect data on households' balance sheets. Data on income are not first priority, but the collection of reliable income data is essential for several analytical purposes. For example, it is useful to analyse indicators on wealth and liabilities by household groups classified by their level of income. Furthermore, indicators on financial vulnerability, such as the debt-income ratio or the debt service-income ratio, are frequently used to assess financial stability of households. The drawback is that it is not possible to comprehensively collect both wealth and income data in a single survey, because it may excessively increase respondent fatigue. Consequently, only gross income is collected in all national datasets of the HFCS. Net income is collected only in a few countries, in some cases from administrative registers.

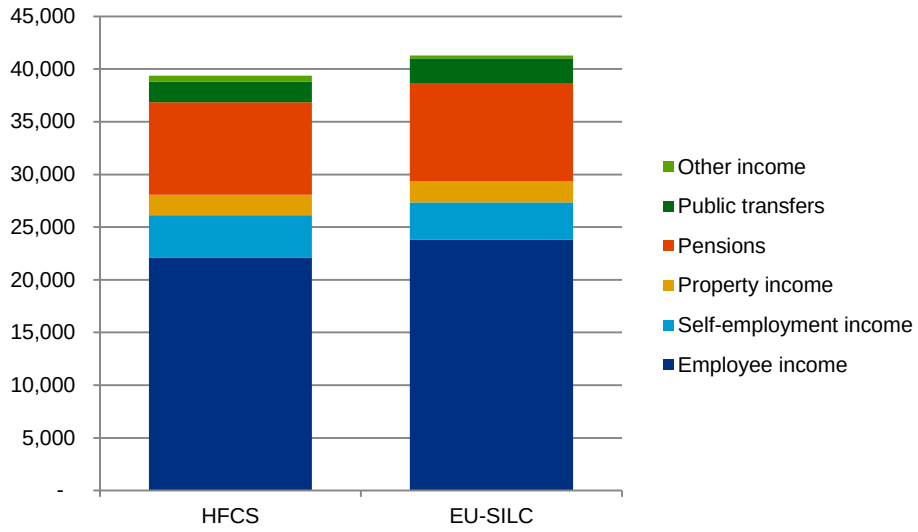
The concept of gross income in the HFCS is identical to the one used in EU-SILC, which is the most complete harmonised survey on household income in Europe. The structure and distribution of gross income can thus be compared between the two data sources.

Figure 10.6 shows the levels of income per household in the euro area, excluding Lithuania[43], produced by the second wave of the HFCS and EU-SILC. Data from EU-SILC are taken from the reference years of the HFCS for each country. In EU-SILC, average gross income per household is €41,300, while it is €39,400 in the HFCS. The difference is €1,900, or 4.8%, which shows relatively good comparability for a survey not specialised in the collection of income. A difference of similar magnitude in euro terms is observed for employee income. The levels of pensions and other social transfers are to a smaller extent higher in EU-SILC. However, the HFCS produces higher (unconditional) averages of self-employment income.

[43] Lithuania did not participate in the HFCS second wave.

Figure 10.6

Structure of gross income in the HFCS and EU-SILC, EUR per household



Sources: ECB – HFCS and Eurostat – EU-SILC.

The differences in the structure of income are in line with expectations. During a wealth survey, there is probably less recall bias for income items related to wealth, namely self-employment and property income. In addition, property income in the HFCS is collected in three questions ("rental income from real estate properties", "income from financial investments" and "income from private business other than self-employment")[44], while the corresponding item in EU-SILC is covered by only two items ("income from rental of a property or land" and "interest, dividends, profit from capital investments in unincorporated business"). Questions on transfer income are however more detailed in EU-SILC than in the HFCS.

Given that the main motivations of collecting income data in the HFCS arise from distributional and vulnerability analysis, it is not only the correct levels of income that matter. The HFCS should also produce a reliable picture of income distribution. The main purpose of the following comparison is to assess the comparability of HFCS income data used in the reporting of the results (Household Finance and Consumption Network, 2016) with EU-SILC. It does not intend to draw a different picture of income distribution than that given by EU-SILC. For this, the HFCS data is not ideal, given the definition of household gross income.

Income distribution statistics (such as EU-SILC) use equivalised household disposable income in measuring inequality, and income is measured at personal rather than at the household level. This income measure is calculated by first assigning the household-level total net income to all household members, regardless

[44] In Italy, income from financial assets and interest paid are estimated by applying to the amounts reported by respondents some relevant return rates derived from external information

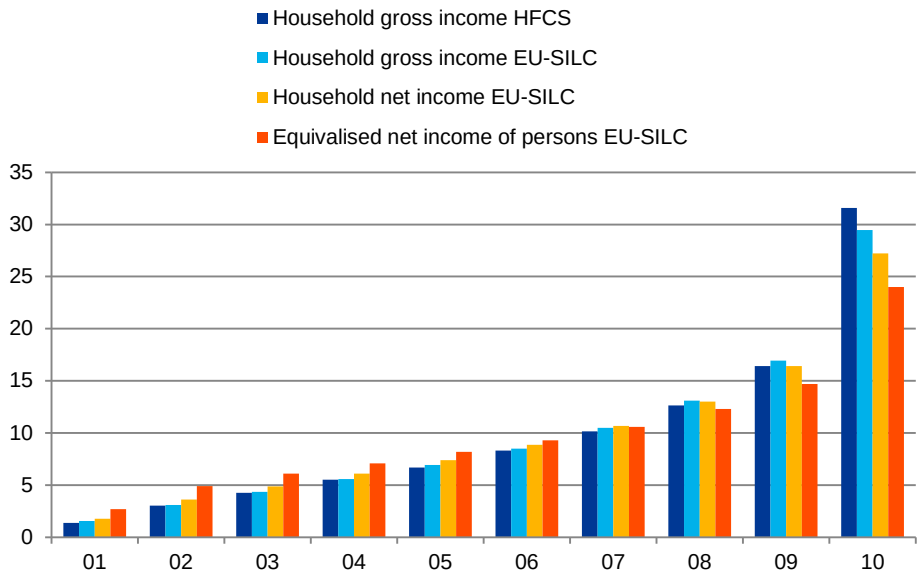
of age, and dividing it by the number of consumption units in the household.[45] Compared with measuring just household-level gross income, this is a better approach for distributional analysis.

The HFCS uses gross income and measures distributions by households. This is consistent with the approach on collecting and measuring wealth information at the household level. Thus, data on distributions in HFCS publications is very different from such data in EU-SILC publications. However, it is possible to calculate income distribution data with definitions identical to the EU-SILC data.

Figure 10.7 shows three different distributions of income derived from EU-SILC data, and compares them with the distribution produced from the HFCS.[46] The columns indicate the share of total income in each income decile. Only comparisons between the distributions of gross income should be used to assess the coherence of HFCS data with EU-SILC data. The other two data series are shown here solely to illustrate how different income concepts result in very different outcomes in the reporting of the HFCS and EU-SILC.

Figure 10.7

Share of total income by income deciles in the HFCS and EU-SILC with various income concepts



Sources: ECB – HFCS and Eurostat – EU-SILC.

The left-most series is that drawn from the HFCS data, which shows the distribution of gross income by households. The second series is produced from EU-SILC data

45 The equivalence scale used in EU-SILC assigns a value of 1 for the first adult member of the household, a value of 0.5 to all other members aged 14 or over, and 0.3 to all members aged 13 or under.

46 The HFCS data, as well as household level data series from EU-SILC are for the euro area, excluding Lithuania, to enable a comparison between identical or similar concepts. The data from EU-SILC on equivalised income is from the reference year 2013 (EU-SILC 2014) for all euro area countries, to enable a comparison with figures published by Eurostat.

with identical definitions. The difference between these two columns indicates the comparability of income data between the two datasets.

Even with identical definitions, the HFCS shows a somewhat more unequal distribution of income than EU-SILC, although the differences are not dramatic. In EU-SILC data, the share of gross income of the bottom 50% of the distribution is 21.5%, while it is 20.9% in the HFCS. The biggest difference is observed in the shares of the top income decile. In EU-SILC data, their share of total household gross income is 29.5%, and in the HFCS 31.6%.

These divergences can be explained by the differences in the structure of income described earlier. HFCS provides higher estimates for self-employment income while EU-SILC provides higher estimates of transfer income, which usually has an equalising impact on income distribution. Additionally, oversampling of wealthy households may have an impact on the share of households with very high income in the HFCS.

The third and fourth data series in Figure 10.7 point to the difference in the income definitions between the HFCS and income distribution statistics. These figures should not be used to assess the coherence of the HFCS data, and are shown to point out why income distribution measures published by other statistics yield very different results to the HFCS.

The third series shows the distribution of net disposable income by households, and the fourth series the distribution of equivalised disposable income of persons. Both data are from EU-SILC, the latter being that available in the public Eurostat EU-SILC database.[47] As expected, net income by households is more equally distributed than gross income. Income taxation is progressive in most European countries, and social transfers are targeted at income-poor households. However, the main difference between the income distribution figures published by the HFCS and EU-SILC comes from shifting from household-level to personal-level equivalised income. The series on the right-hand side show the distribution of income.

To conclude, the level, structure and distribution of household gross income produced by the HFCS is fairly coherent with the corresponding information produced by EU-SILC. However, the concepts and methodologies used in the reporting of the results of the two statistics are very different and are not comparable.

[47] http://ec.europa.eu/eurostat/web/income-and-living-conditions/data/database, EU-SILC figures for equivalised disposable income are for the whole euro area.

Appendices

HFCS definitions of financially knowledgeable person and HFCS household definition

HFCS definition of Financially knowledgeable person (FKP)

The Financially knowledgeable person (FKP) is defined as the person who is most knowledgeable on financial matters regarding both the household as a whole and its individual members. He/she will be invited to provide a large part of the information requested during the interview.

HFCS Household definition

The target reference population for national surveys is all private households and their current members residing in the national territory at the time of data collection. Persons living in collective households and in institutions are generally excluded from the target population.

Household is defined as a person living alone or a group of people who live together in the same private dwelling and share expenditures, including the joint provision of living essentials. Employees of other residents (i.e. live-in domestic servants, au-pairs, etc.) and flatmates without other family or partnership attachments to household members (e.g. resident boarders, lodgers, tenants, visitors, etc.) are considered separate households.

Subject to the further and specific conditions shown below, the following persons must, if they share household expenses, be regarded as household members:

1. persons usually resident, related to other members
2. persons usually resident, not related to other members
3. persons usually resident, but temporarily absent from dwelling (for reasons of holiday travel, work, education or similar)
4. children of the household being educated away from home
5. persons absent for long periods, but having household ties: persons working away from home
6. persons temporarily absent but having household ties: persons in hospital, nursing home, boarding school or other institution

Further conditions for inclusion as household members are as follows:

for persons usually resident, but temporarily absent from the dwelling (3):

- the person currently has no private address elsewhere and the actual or intended duration of absence from the households is less than six months

for children of the household being educated away from home (4) and persons absent for long periods, but having household ties, such as persons working away from home (5):

- irrespective of the actual or intended duration of absence, if the person is the partner or child of a household member, continues to retain close ties with the household, regularly returns to this address (for instance, at the end of the academic term) and considers it to be his/her main residence.[48]

for persons temporarily absent but having household ties: persons in hospital, nursing home, boarding school or other institution (6):

- the person has clear financial ties to the household and the actual or expected duration of absence from the household is less than six months

Sharing in household expenses includes benefiting from expenses (e.g. children, persons with no income) as well as contributing to expenses. If expenses are not shared, then the person constitutes a separate household at the same address.

A person will be considered a usually resident member of the household if he/she spends most of his/her daily night-rest there, evaluated over the past six months (this includes children in joint custody and elderly parents if they spend more days living in the household dwelling than anywhere else).

Persons forming new households or joining existing households will normally be considered members at their new location; similarly, those leaving to live elsewhere will no longer be considered members of the original household. The above mentioned “past six months” criteria will be replaced by the intention to stay for a period of six months or more at the new place of residence. Account has to be taken of what may be considered as “permanent” movements in or out of households. Thus a person who has moved into a household for an indefinite period or with the intention to stay for a period of six months or more will be considered a household member, even though the person has not yet stayed in the household for six months, and has in fact spent a majority of that time at some other place of residence. Similarly, a person who has moved out of the household to some other place of residence with the intention to stay away for six months or more will no longer be considered a member of the previous household.

If the person who is temporarily absent is in private accommodation, then whether they are members of this (or their other) household depends on the length of their absence.

[48] The definition of household membership differs slightly in Italy, as it includes persons in cases (4) and (5) as members of the households only if they are absent for less than six months.

Exceptionally, certain categories of persons with very close ties to the household may be included as members irrespective of the length of absence, provided they are not considered members of another private household. In particular, students that live elsewhere but retain close ties with the household, regularly return to this address and consider this address to be their main residence are to be considered part of the household irrespective of their length of stay at the other address.

Coverage issues: in the application of these criteria, the underlying intention should be to minimise the risk that individuals who have two private addresses at which they might potentially be enumerated are not double-counted in the sampling frame. Similarly, the intention should be to minimise the risk of some persons being excluded from membership of any household, even though in reality they belong to the private household sector.

Persons living in collective households and the institutionalised population are excluded from the survey population and not covered:

Collective household: refers to a non-institutional collective dwelling such as a boarding house, dormitory in an educational establishment or other living quarters shared by more than five persons without sharing household expenses. Also included are persons living as lodgers in households with more than five lodgers.

Institution: refers to old persons' homes, health care institutions, religious institutions (convents, monasteries), and correctional and penal institutions. Basically, institutions are distinguished from collective households, in that in the former, the resident persons have no individual responsibility for their housekeeping. In some cases, old persons' home can be considered collective households on the basis of this last rule.

Changes in the core variables for wave 2

As mentioned in Chapter 2, there are several new variables collected in the second wave of the HFCS, as well as some changes in definitions. The list of the most important changes in output variables is shown in Table A.1.

Table A.1

Changes in the core variables for wave 2

Real assets and their financing	
HB113$x HMR mortgage: reason for refinancing	New variable
HB115$x HMR mortgage: mortgage renegotiation	New variable
HB260$x Other property use	Four questions (HB260$x-HB263$x) on the use of other properties were modified to one multiple choice question with six answer categories
HB260$x Other property: for household own use	Definition and answer categories changed, HB260$x=1 has the same interpretation in both waves
HB261$x Other property: for business use	Variable removed, HB260$x=2 in the second wave corresponds HB261$x=1 in the first wave
HB262$x Other property: leased or rented	Variable removed, HB260$x=3 in the second wave corresponds HB262$x=1 in the first wave
HB263$x Other property: for other use	Variable removed, HB260$x=4 in the second wave corresponds HB263$x=1 in the first wave, HB260$x=5 in the second wave corresponds to HB263$x=2 in the first wave and HB260$x=6 in the second wave corresponds to HB263=3 in the first wave.
HB313$x Other property mortgage: reason for refinancing	New variable
HB315$x Other property mortgage: loan renegotiation	New variable
HB4800 Purchase of vehicles	New variable
HB4810 Price of purchased vehicles	New variable
Other liabilities, credit constraints	
HC0330 Has private loans	New variable
HC0340 How many private loans	New variable
HC035$x Private loan: purpose of loan	New variable
HC036$x Private loan	New variable
HC0370 Additional private loans, outstanding amount	New variable
HC0400 Has any non-collateralised loans, HC0410 number of non-collateralised loans, HC050$x non-collateralised loan $x: purpose of the loan, HC060$x non-collateralised loan $x: amount initially borrowed, HC070$x non-collateralised loan $x: initial length of the loan, HC080$x non-collateralised loan $x: outstanding balance of loan, HC090$x non-collateralised loan $x: current interest rate of loan, HC100$x non-collateralised loan $x: monthly payment on loan, HC1100 total amount owed for additional non-collateralised loans, HC1200 monthly payment on additional non-collateralised loans	Private, informal loans from relatives and friends excluded from question wording and/or survey definition
Private businesses, financial assets	
HD1320g Market value of mutual funds – aggregate amount all funds together	Change of definition, refers to values of unknown mutual fund types. Information included in variable HD1330 in the second wave.
HD1330 Market value of mutual funds – all funds together	Replacing variable HD1320g, to be completed only if values by types of mutual funds cannot be provided by the respondents
Consumption and saving	
HI0100 Amount spent on food at home	Reference period and question wording clarified as a typical month instead of typical month in the past 12 months.
HI0200 Amount spent on food outside home	Reference period and question wording clarified as a typical month instead of typical month in the past 12 months.
HI0210 Amount spent on utilities	New variable
HI0220 Amount spent on consumer goods and services	New variable

Coverage of the core items in the second wave of the HFCS

The surveys of countries that started new surveys in the first or second HFCS wave or replaced their previous survey with the new HFCS to supply the HFCS data are largely built on the euro area blueprint questionnaire. Some countries (Italy, Spain, Finland, France, the Netherlands) have adapted existing national surveys to the HFCS. In the countries with pre-existing surveys, harmonisation may involve some approximation until the survey gradually converges to the output variables of the HFCS. The whole content of the HFCS blueprint questionnaire may therefore not always be fully covered in all countries in the first survey waves. Box A.1 provides information on the incomplete coverage of the HFCS core questions in the second wave of the HFCS.

In addition to the core items available in the user database for external researchers, the HFCS collects confidential information from all households selected in the sample. Details on the contents and country coverage of this so-called sample register file are provided in Box A.2.

Box A.1
HFCS core variables not covered in HFCS wave 2

Demographics

Questions on country of birth and length of stay in the country are not collected in Spain. In France and the Netherlands, these variables are not collected for foreign-born persons.

Real assets and their financing

The question on how long the household has lived in the current household main residence (HMR) is not asked in Finland. The questions on the amount of rent paid for partially-owned household main residence and on the percentage of ownership of the household main residence are not asked in the Netherlands. The questions on the year the household main residence was acquired and its value at the time of acquisition are not asked in Finland, and neither is the question on the percentage of ownership of other properties.

The questions on the number of cars and the number vehicles other than cars are not asked in Italy. The question on the number of vehicles other than cars is only partially asked in Spain and France. The questions on the ownership and value of valuables (such as jewellery, works of art, antiques) are not asked in Finland. The question on purchase of vehicles is not asked in Finland or Poland.

Questions on both HMR and other property mortgage refinancing, reasons for refinancing, year when loan was taken or refinanced, initial amount borrowed, length of loan, adjustable interest rate and current interest rate are not asked in Finland. Questions on reasons for mortgage refinancing and on mortgage renegotiation are not asked in Italy or Poland. Questions on the current interest rate of mortgages are not asked in Poland. The question on additional borrowing on HMR and other property mortgages is not asked in Italy, and the question on additional borrowing on other property mortgages is not asked in Germany or Spain. Questions on other property mortgage refinancing

and renegotiation, and the reasons for refinancing are not asked in Spain. Questions on the adjustable interest rate are not included in the data from Malta. In France, mortgages collateralised by properties are defined either as mortgages or as loans that are taken for the purpose of buying the property and that have an insurance scheme (*sociétés de cautionnement*) – a type of guarantee.

Other liabilities, credit constraints

The question block on private loans is not asked in Estonia, France or Poland. In Finland, questions on the number and purpose of private loans are not asked, and the outstanding amount on private loans is given as the outstanding amount on additional private loans. Questions on the amount initially borrowed, the initial length and current interest rate of non-collateralised loans are not asked in Finland.

Data on monthly payments on non-collateralised loans are only partially collected in Germany and Italy. The question on re-applying for credit after refusal is not asked in Spain or Italy.

Private businesses, financial assets

The question on household members working in self-employment businesses is not asked in Finland or Poland. The question on the percentage of self-employment business ownership is not asked in Finland.

The value of saving accounts is not separately collected in Poland, and is provided jointly with data on sight accounts. Sub-items of the mutual funds questions are not separately collected in Finland, and values by type of mutual fund are not separately collected in Portugal. Sub-items on mutual funds predominantly investing in real estate and hedge funds are not separately collected in Italy. The sub-item on mutual funds predominantly investing in hedge funds is not separately collected in Spain or France. The question on types of bonds owned is only partially collected in Finland and Poland. The question on foreign shares in the owned shares' portfolio is not asked in France or Spain. The questions on money owed to the household and extra assets in managed accounts are not asked in Finland. The value of any other assets is only partially collected in Spain.

Employment

The secondary labour status question (in addition to the main labour status) is not asked in Finland or the Netherlands. Information on the type of contract is not collected in Finland. Data on time spent in the current main job is not collected in Finland and collected only for employees in Spain. Information on the type of secondary employment is not collected in Finland or Poland. The question on total time spent in employment since the age of 16 and expected retirement age is not asked in Finland. Data on expected retirement age is only partially collected in Italy and Spain.

Pensions and life insurance policies

The variables on public pension plans are not collected in Finland. The variable on the current percentage of gross earnings contributing to public pension plans is not collected in Austria, Italy, Spain or Poland, and is collected only for the main plan in Belgium and Luxembourg. The current value of accounts in public pension schemes is not collected in Austria, Spain, France, Greece, Hungary, Italy, Luxembourg, Malta or Poland.

The questions on occupational pension plans are not asked in Hungary. The current value of occupational pension schemes that have an account is not collected in Greece. The value of whole life insurance contracts is not collected in Finland. The question on whether the occupational plan has regular benefits in retirement is not asked in Finland, Greece or Poland. The question on the type of voluntary public pension plan (pension plan / whole life insurance contract) is not asked in Austria.

Income

Received income from private businesses other than self-employment is not collected in Italy. Income from private and occupational pension plans is not separately collected in Spain, but rather provided together with income from public pension plans. The question on the character of collected annual income (higher/normal/lower) and the question on future income expectations are not asked in Ireland or Finland.

Intergenerational transfers, gifts

In Finland, only the question on whether a substantial gift/inheritance was received is asked, the rest of the gift/inheritance block is not collected. Questions on from whom the gift/inheritance was received and whether the household expects to receive a gift/inheritance are not asked in Spain.

Consumption and saving

The amount spent on food outside the home is not collected separately, but rather provided together with amount spent for food at home in Spain. The amount spent on consumer goods and services is not collected in Poland. The question on the purpose of saving is only partially asked in Italy. Questions on the character of the last 12 months' expenses (high/normal/lower) and on the comparison of the balance between income and expenses is not asked in Finland. The question on the source of extra income to meet expenses in households with expenses above income is not asked in Finland, and is only partially collected in Spain. The question on the ability to get financial assistance from friends or relatives is not asked in Ireland or Spain.

Box A.2
The sample register file

Purpose and contents

In addition to the core and non-core variables that are provided in the user database for external researchers, the HFCS collects information from all households initially selected for the sample in a so-called sample register file (S-file). The S-file contains data on the mode and outcome of the interview, interview dates and numbers of contacts, interviewers' assessments of the dwelling as well as of the neighbourhood, strata, sampling units and case design weights. This information is not provided for research use, mainly due to confidentiality reasons. Detailed information on dwellings and their surroundings together with data on sample units, often linked to geographical information, increases the identification risk of households significantly. However, such data provide important tools for analysing survey non-response. Variables in the S-file together with e.g. data on interviewers' and regional information available only at the national level are often used in estimating response propensities during the construction of survey weights (Albacete et al., 2016).

Availability of sample register information in the HFCS second wave

Due to their importance in non-response analysis, variables in the sample register file are flagged as core variables, i.e. they should be collected in all countries participating in the survey. However, the availability of such information varies significantly between countries. In some countries, these data cannot be disseminated because of confidentiality reasons. Additionally, few countries are able to provide these data only for respondents. For example, detailed information on dwellings is not available in the Netherlands or in Finland[49], where the survey mode is either CATI or CAWI. Availability of sample register information is shown in Table A.2. This table indicates which variables are available for centralised evaluation at the ECB. There is no distinction between information not collected and information not disseminated due to confidentiality reasons.

Table A.2
Availability of variables in the sample register file

Country	Non-respondent included	Strata and design weight	Number of variables on dwelling assessment	Information on contacts
Belgium	Yes	Yes	7	Yes
Germany	No	No	1*	No
Estonia	Yes	Yes	7	Yes
Ireland	No	No	0	No
Greece	Yes	Yes	4*	Yes
Spain	Yes	No	6	Yes
France	Yes	No	3	Yes
Italy	Yes	Yes	6	Yes
Cyprus	Yes	Yes	6*	Yes
Latvia	Yes	Yes	7*	Yes
Luxembourg	No	No	1*	No
Hungary	Yes	Yes	7	Yes
Malta	Yes	Yes	7	Yes
Netherlands	Yes	No	0	No
Austria	Yes	Yes	7	Yes
Poland	Yes	Yes	7	Yes
Portugal	Yes	No	4	Yes
Slovenia	Yes	Yes	7	Yes
Slovakia	Yes	Yes	7	Yes
Finland	Yes	Yes	1	Yes

Source: HFCS metadata.
* only for respondents

Collection of the non-core items

The following table provides an overview of non-core variables covered in one or more of the HFCS country files in wave 2.

[49] Finland collects information on the type of dwelling from administrative sources.

Table A.3

HFCS non-core variables collected in national surveys

Demographics	
RNA0200 Citizenship	France, Italy, Luxembourg
PNA0100 Field of study	Spain, Italy
PNA0200 Health	Spain
PNA0300 Siblings	France
PNA0400 Are you the eldest	France
PNA0500 RP's/partner's father alive	France
PNA0501 RP's/partner's mother alive	France
PNA0510x Age of father and mother	France
PNA0600x Education of father/mother	Italy, Portugal
PNA0700 Occupation of father	Spain, France, Portugal
PNA0701 Occupation of mother	Spain, France, Portugal
PNA0850 Legal arrangements for marriage or recognised partnership	Spain, France
PNA0851 Sort of legal arrangement for marriage or recognised partnership	Spain, France
Real assets and their financing	
HNB0800 HMR/any part used for business purposes?	France
HNB0810 HMR – year of construction	Greece, Spain, Italy, Portugal, Finland
HNB0910x HMR – External support for housing acquisition	Greece , France, Luxembourg, Portugal
HNB0920 HMR/Imputed rent	Italy, Finland
HNB1000 House prices expectations	Slovakia
HNB1120 Year of most substantial maintenance	Luxembourg
HNB1150 Expected price of your home	Greece
HNB130$x HMR mortgage1: institution you have loan with	Spain, France
HNB140$x HMR mortgages: work for institution granting the loan	Portugal
HNB1700 Overpaying/voluntary step-up payments on HMR mortgages	Portugal
HNB1710 Monthly amount of extra voluntary payments on HMR mortgages	Portugal
HNB1800 Rent net or including other charges	France
HNB190$x Other property: how property was acquired	Spain, France, Italy
HNB2000 Remaining other properties: renting out of property	France, Italy
HNB2010 Other properties: how much rent is collected	Italy
HNB2300 Overpaying/voluntary step-up payments: loans on other properties	Portugal
HNB2310 Monthly amount of voluntary payments: loans on properties other than HMR	Portugal
HNB2800 Sold properties or consumer durables	Spain
HNB2820 Amount received – sale of properties or and consumer durables	Spain
HNB3000 Reasons for moving	Portugal

Other liabilities, credit constraints

HNC004$x non-collateralised loan: year the loan was taken	France, Portugal
HNC005$x non-collateralised loan: nature of the lender	Spain, France
HNC0125 Late or missed payments on loans	Spain, France, Italy, Portugal
HNC0126 Any outstanding overdue payments	Portugal
HNC0127 Any overdue payments by more than 90 days	Portugal
HNC0200x Reasons for being refused credit	Spain, France

Private businesses, financial assets

HND010$x Business: year the business was started	Spain, France, Portugal
HND020$x Business: last year's total business sales	France, Portugal
HND0400 Any guarantees provided to businesses	Spain, Portugal
HND0410 Value of the guarantees provided to businesses	Spain
HND0420 Any guarantees provided to non-HH members	Portugal
HND0600 Is interest paid on sight accounts	Spain
HND0800 Are all accounts in euro	Portugal
HND1000x Market value by type of bond	Italy
HND1800 Number of different shares (companies)	Spain
HND1900 Any shares in company you work for	Spain
HND1910 Value of shares of the employer company	Spain
HND2200 Assets deposited abroad	Portugal
HND3000x Largest assets in HH balance sheet	Belgium
HND3010 Portfolio shifts last two years?	Belgium
HND3020 Portfolio shifts last two years: money out	Belgium
HND3030 Portfolio shifts last two years: money in	Belgium
HND3040 Would not invest again?	Belgium
HND3050x Assets HH would not invest again	Belgium
HND3100 Net worth past two years	Belgium, France, Portugal
HND3200 Net worth next two years	Belgium, Portugal
HNF0100x Has other insurance policies (accidents, theft, fire etc.)	Spain

Employment

PNE0100 Seasonal employment	Italy, Portugal, Slovakia
PNE0110 Number of working weeks per year	Italy, Slovakia
PNE0200 Gross monthly income – main job (employees)	Spain, Portugal, Slovakia
PNE0300 Gross monthly income from self-employment	Spain, Portugal, Slovakia
PNE0500 Private-public organization	France, Italy, Luxembourg, Portugal, Finland
PNE0600 Number of employees – main employer	Spain, Portugal
PNE0700 Hours worked – additional employment contracts (as an employee)	Spain, Italy, Slovakia
PNE0800 Gross monthly income from additional jobs	Spain, Slovakia
PNE0900 Probability of losing job	Greece, Spain, Slovakia
PNE1000 Looking for job	Spain, France, Greece, Slovakia
PNE1100 Expect find new job in next 12 months	Greece, Spain, Slovakia
PNE1300 Hours a week would like to work in new job	Slovakia
PNE1400 For what minimum wage would work	Spain, Slovakia
PNE1600 Year they stopped being employed (for retirees)	Portugal
PNE1700 Employment status in last main job	France
PNE1800 Full time/part time – last job	France

PNE1900 What did firm/organisation you worked for make or do	Spain, France
PNE2000 Former job title and description / ISCO	Spain, France, Portugal
PNE2100 Time in former employment	France
PNE2200 Total time in full-time employment	Spain, France, Luxembourg
PNE2210 Total time in all part-time employment	Spain, France
PNE2300 Prevailing employment situation in working life	Spain
PNE2400 No of different employers	Spain
PNE2500 Longest time with one employer	Spain
PNE2700x Worsening of job conditions past 2 years	Greece, France, Portugal
PNE2800x Expected worsening of job conditions next 2 years	Portugal

Pensions and life insurance

PNF0401 Public pension plan: years contributing	France, Italy
PNF0501 Public pension plan: expected age to receive benefits	Italy
PNF070$x Occupational plans: name	France
PNF0720 Current value of all occupational plans that do not have an account.	Netherlands, Finland
PNF100$x Occupational pension plan: is employer contributing	Spain, France, Italy
PNF120$x Occupational pension plan: years contributing	Spain, France, Italy, Slovakia
PNF130$x Occupational pension plan: plan has an account balance	Spain, France, Italy
PNF131$x Occupational pension plan: value of account	Spain, France, Italy
PNF160$x Occupational plans: payment to be received at retirement age	Spain
PNF180$x Occupational pension plan: expected age of collecting pension	Spain, France, Italy
PNF2000 Number of voluntary private pension plans	Spain, France, Italy
PNF210$x Voluntary pension plan: type of voluntary pension plan	Spain, France, Italy
PNF220$x Voluntary pension plan: years contributing	Spain, France, Italy, Slovakia
PNF230$x Voluntary pension plan: contributions	Spain, France, Italy
PNF240$x Voluntary pension plans: how contributions are set	France
PNF280$x Voluntary pension plan: age to start receiving payments	France, Italy
PNF290$x Voluntary pension plan: kind of payment at retirement age	Spain, France
PNF300$x Voluntary pension plan: value of account	Spain, France, Italy
PNF310$x Voluntary pension plan: whole life insurance policy: cash value	Spain, France
PNF311$x Voluntary pension plan: expected age to collect pension	Spain, France, Italy
PNF3600 Has private health insurance	Spain, Italy
PNF3610 Monthly payments for health insurance policy(ies)	Italy
PNF3700 Expectations about pension income	Slovakia

Income

HNG0110 Net income from regular social transfers	Italy, Poland
HNG0210 Net income from regular private transfers	Italy, Poland
HNG0310 Net rental income from real estate property	Italy, Poland
HNG0410 Net income from financial investments	Italy, Poland
HNG0510 Net income from private business other than self-employment	Italy, Poland
HNG0610 Net income from other sources	Italy, Poland
HNG0710 Income taxes and social contributions	Italy, Finland
PNG0110 Net employee income	Italy, Poland

PNG0210 Net self-employment income	Italy, Poland
PNG0310 Net income from public pensions	Italy, Poland
PNG0410 Net income from private and occupation pension plans	Italy, Poland
PNG0510 Net income from unemployment benefits	Italy, Poland
Intergenerational transfers, gifts	
HNH0500 Substantial gift made to children/other people outside household	France, Portugal
HNH0600 Who was the beneficiary of the gift	France, Portugal
HNH0700 Year donation was made	France, Portugal
HNH0800 How much was donation made worth	France, Portugal
Consumption and saving	
HNI0200 Meet any regular payments	Portugal, Slovakia
HNI0210 Expenditure on regular payments	Portugal, Slovakia
HNI0500 Comparison of future expenses with current level	Spain
HNI0700 More or less savings in the next year	Belgium, Slovakia
HNI0800 General price expectations	Belgium, Luxembourg, Slovakia
HNI1000 General personal financial situation expectations	France, Slovakia
Payment habits and financial literacy (non-core section)	
HNJ1100 Any debit or/and ATM cards	Spain, Italy
HNJ1200 How frequently uses debit card	Spain
HNJ1300 Frequency of cash withdrawals in ATMs	Spain
HNJ1400 Use of direct debit	Spain
HNJ1500x Type of payments by direct debit	Spain
HNJ1600x Reasons for not using direct debit	Spain
HNJ1800 Payments by bank cheques	Spain
HNJ2000 Any payments received by credit transfer	Spain
HNJ2300a No of credit cards	Italy
HNJ2800x Ever used other means of payment	Spain
HNJ2900 Link used for info or payments	Spain
HNJ3100 A computer at home	Finland
HNJ3200 Any household member use the internet	Spain
HNJ3800 Cash at home to meet normal needs	Spain, Hungary, Slovakia
HNK0400 General economic situation expectations	Slovakia
HNM0100 Financial literacy Variable/fixed interest rates	Luxembourg, Slovakia
HNM0200 Financial literacy Inflation	Luxembourg, Slovakia
HNM0300 Financial literacy Portfolio diversification	Luxembourg, Slovakia
HNM0400 Financial literacy Riskiness	Luxembourg, Slovakia

Statistical disclosure: additional information

The variable identifiers below refer to the variable names in the User Database (UDB).

Table A.4

Variables added, deleted, or recoded in the UDB

Sample register file

The sample register file is not provided in the User Database. The following variables are recoded into the User Database.

SB010$x, SB030$x (date and time of interview)	Recoded to SB1000 (quarter or year of interview), e.g. 2009Q2 or 2010.
All other variables in the S file	Dropped, except: * SA0100 country * SA0200 survey vintage * SA0110 past household ID (panel) * SA0210 vintage of past interview which are copied without modification to H file

Demographics

RA0300 age	Top-coded at 85. Applied in all countries
RA0300_B age (brackets)	Created "age, coded in 5-year brackets" from RA0300 [0,5), [5,10),…[80,85), [85,+∞). Applied in IE and MT.
RA0400 country of birth	Recoded to "local"/OEA/OEU/OTH where local is the ISO two digit code for the survey country[50] Applied in all countries.
PA0100 marital status	Consensual union (3) merged into married (1) category. Applied in LU, but code (3) is not available in EE, FI, IE, IT, LV, MT, PL, PT.
PA0200 education	Merge code 4 into code 3, and code 6 into code 5. Applied in all countries.

Real assets

HB0100_B size of HMR	Bracketed version created from HB0100: [0;30), [30;40), [40;50), [50;60), [60;80), [80;100), [100;120), [120;150), [150;200), [200;+∞). Applied in MT.
HB0200 length in residence	Top-coded at 85.
HB0700 year of acquisition	Bottom-coded at 1925.[51]

Financial assets

HD050$x_B number of employees	Created from HD050$x. Brackets: 0, [1;2], [3;9], [10,+∞). Applied in BE, DE, EE, FR, LU, MT and PT.
HD1910 specification of assets	Deleted (verbatim answer)[52]

Employment

PE1000 time in main job	Top-coded at 73. [43]

Pensions

PF0300 years contributing	Top-code at 73. [43]

Income

HG0620 sources of other income	Deleted (verbatim answer) [44]

50 In the case of panel households, this value will not be modified in the event that the country of birth joins the EU or euro area at a later stage.

51 The following variables are either top- or bottom-coded, following the top-coding of age: HB0700 (HMR: year of acquisition), PE1000 (time in main job), PF0300 (pension: years contributing) HH020$x (inheritance: year received). The year of acquisition of the HMR has to be bottom-coded, at 1925 (i.e., 2014-85). If not, the age of a respondent over 85 who has been living in the same house could be deducted (or at least, a new lower bound). 0.1% of households are affected by this cut-off (IT and NL data). The year an inheritance has been received also has to be bottom-coded in the same way. The time in main job and the number of years someone has contributed to the pension have to be top-coded for the rare case that somebody over 85 has continued working. The top-coding assumes that the person started working at age 16, hence the top-coding is at 85-16=69. However, since the minimum school leaving age in 1941 was presumably below 15 in European countries, the top-code could be increased, up to 73 for a school leaving age of 12 in 1941.

52 Verbatim answers may contain identifying information and cannot be protected adequately. They are to be used mostly during the editing phase of the data, and will feed back into questionnaire design.

Inheritances	
HH020$x year received	Bottom-code at 1925. [43]
Interview closure	
HP0100 items difficult	Deleted
HP0200 items missed	Deleted
HP0300 interviewee additions	Deleted
Paradata	
HR0100 to HR1100	Deleted
HR1300 to HR1600	Deleted
Non-core files	
RNA0200 citizenship	Recoded to "local"/OEA/OEU/OTH50. Applied in all countries providing RNA0200.
PNA0600x Education of parents	Merge code 4 in code 3, and code 6 in code 5. Applied in all countries providing PNA0600x.
Top-coding and grouping	
HB2410 number of properties	Top-coded at 5 in EE. Top-coded at 6 in LU and MT. Top-coded at 15 in PT.
HB250x type of other properties	Categories Office, Hotel, Farm and Industrial buildings/Warehouse merged with Other in MT.
HB4310 number of cars	Top-coded at 5 in EE. Top-coded at 4 in BE, LU, MT and PT.
HB4510x number of other vehicles	Top-coded at 2 in EE. Top-coded at 4 for HB4510a (motorbikes) in BE and PT, at 1 for HB4510b (trucks) in BE and PT, at 3 for HB4510c (vans) in PT and 1 in BE, at 3 for HB4510f (other) in PT.
HC0410 number of non-collateralised loans	Top-coded at 5 in BE and PT.
HD0210 number of businesses owned	Top-coded at 2 in BE and EE.
PE0300 occupation	Recoded from 2- to 1-digit ISCO88 codes in GR, LU and MT for disclosure reasons.
PE0400 main employment sector	NACE sectors B to E, L to N, and R to U, each merged in PT.

Revisions to data from the first wave

Some countries have revised the datasets from the first wave. This means that some indicators presented for the first wave in the publications of the second wave may be different from the corresponding indicators published in 2013. In all cases, the need for revisions is based on improvements in the data production methodologies or data sources available.

In Italy, the variables on voluntary pension schemes were revised. This was due to a small error detected in the first-wave data. The revision only had a small impact on the participation rates and average values of voluntary pension schemes.

In Malta, some first-wave variables for panel households were revised based on the information received in the second-wave interviews. Some households with self-employed members did not report any values for self-employment businesses in the first-wave interview. For panel households, these values were estimated based on the answers for the second wave. This led to a small increase in the share of households with self-employment businesses and to a decrease in their mean and median values. Liabilities of panel households that were reported to be taken before

2010 were included also in the first-wave data. Additionally, some variables that were imputed in the first wave, but collected in the second wave were corrected, if the collected and imputed values were very different.

In the Netherlands, missing observations of the value of accounts in voluntary pension schemes were re-imputed with an improved imputation methodology. This led to a slight decrease in participation rates in voluntary pension schemes, and a slight increase in the conditional mean of voluntary pension accounts. The amount still owed for additional mortgages on the household main residence were corrected, due to an error detected in data processing. Due to a small number of households having more than three mortgages on the HMR, the impact of this correction was relatively small.

In Portugal, the weights and their replicates for wave 1 were recalculated using the revised series for the population statistics, which became available after the release of the Census in 2011. Additionally, the model used in the calibration was changed to be in line with the one used in wave 2.

In Finland, the value of household main residence is estimated using transaction price data. The method of valuation was changed for the second wave, and consequently the first-wave data were revised to maintain comparability between the waves. The new method uses different (nominal instead of real) and regionally more detailed price indices to uprate the original registered purchase prices of housing co-operatives to the reference year level. For detached houses, the estimation method was also adjusted to better account for the year of construction. The impact of this change was significant for values of housing co-operatives. The total HMR value increased by 4.3%, and the share of households with negative net housing wealth decreased. These revisions had a particularly large impact on households in the middle of the distribution and consequently on median net wealth that increased due to this revision. Data on other properties were also revised. First of all, the identification of free-time residences was revised, resulting in fewer households with multiple free-time residences. The value of farm land was estimated on the basis of tax values in the original data. In the revised data, values are market values estimated on the basis of transaction prices.

Additionally, in Finland, the market values of mutual funds were multiplied by a factor of 1/0.7, since the previous data for the first wave consisted of tax values of 70% instead of market values. Finally, for investments in businesses not publicly traded, the second-wave data cover all legal forms and not just limited liability companies as in the original first-wave data. For the first wave, the net value of partnerships was available from tax data and added to the value of self-employment businesses in the revised data. Even after this correction, the first- and second-wave data are not fully comparable because the first-wave data does not cover sole proprietorships. However, based on the second-wave data, net value of limited liability companies and partnerships accounts for 95% of the total net value of self-employment businesses in Finland.

References

Albacete, N., Lindner, P. and Wagner, K. (2016), "Eurosystem Household Finance and Consumption Survey 2014. Methodological notes for Austria", *Addendum to Monetary Policy & the Economy* Q2/16.

Altissimo, F., Georgiou, E., Sastre, T., Valderrama, M., Sterne, G., Stocker, M., Weth M., Whelan K., and Wilman, A. (2005), "Wealth and asset price effects on economic activity", *ECB Occasional Paper* No 29.

American Association for Public Opinion Research (AAPOR) (2011), "Standard definitions. Final dispositions of case codes and outcome rates for surveys", AAPOR.

Andreasch, M. and Lindner, P. (2014), "Micro and macro data: A comparison of the Household Finance and Consumption Survey with financial accounts in Austria", *ECB Working Paper Series* No 1673, May.

Barceló, C. (2006), "Imputation of the 2002 wave of the Spanish survey on household finances (EFF)", *Documentos Ocasionales*, No 0603, Banco de España.

Barnard, J. and Rubin, D.B. (1999), "Small-sample degrees of freedom with multiple imputation", *Biometrika,* Vol. 86, Issue 4, pp. 948-955.

Berben, R.P., Bernoth, K., and Mastrogiacomo, M. (2006), "Households' response to wealth changes: do gains or losses make a difference?", *DNB Working Paper* No 90.

Berger, Y.G. (2004), "Variance estimation for measures of change in probability sampling", *Canadian Journal of Statistics*, Vol. 32.

Berger, Y.G. and Priam, R. (2010), "Estimation of correlations between cross-sectional estimates from repeated surveys – an application to the variance of change", *Proceedings of the 2010 Statistics Canada Symposium*.

Biancotti, C., Kennickell, A. B., Sánchez Muñoz, C. (2009), "Decentralised multi-country imputation: the case of the euro area Household Finance and Consumption Survey", *Supporting Paper* No 25, UNECE Conference of European Statisticians.

Bledsoe, R., and Fries, G. (2002), "Editing the 2001 survey of consumer finances", *Annual Meeting of the American Statistical Association, Joint Statistical Meetings*, New York, New York, 11-15 August 2002.

Browning, M., Crossley, T.F., and Weber G. (2003), "Asking consumption questions in general purpose surveys", *The Economic Journal*, Vol. 113, No 491, Features (Nov. 2003), pp. F540-F567.

De Leeuw, E. D. (2005), "To mix or not to mix data collection modes in surveys", *Journal of Official Statistics*, Vol. 21, No 2, pp. 233-255.

Dillman, D.A. and Christian, L.M. (2005), “Survey mode as a source of instability in response across surveys”, *Field Methods*, Vol. 17, Issue 1, pp. 30-52.

Elamir, E.A. H and Skinner, C.J. (2006), “Record level measures of disclosure risk for survey microdata”, *Journal of Official Statistics*, Vol. 22, Issue 3, pp. 525-539.

Engelhardt, G.V. (1996), “House prices and home owner saving behavior”, *Regional Science and Urban Economics*, Vol. 26, No 3/4, pp. 313-336.

Eurostat (2011a), “EU-SILC, Description of target variables: cross-sectional and longitudinal”, European Commission, Eurostat.

Eurostat (2011b), “EU legislation on the 2011 population and housing censuses, explanatory notes”, European Commission, Eurostat.

Eurostat (2013), “Handbook on precision requirements and variance estimation for ESS households surveys”, Eurostat.

UNECE (2011), *The Canberra Group Handbook on Household Income Statistics, second edition*.

Faiella, I. (2010), “The use of survey weights in regression analysis”, *Working Paper* No 739, Banca d’Italia.

Girard, C. (2009), “The Rao-Wu rescaling bootstrap: from theory to practice”, *Federal Committee on Statistical Methodology Research Conference, 2009*.

Groves, R.M. and Peytcheva, E. (2008), “The impact of non-response rates on non-response bias”, *Public Opinion Quarterly*, Vol. 72, Issue 2, pp. 167-189.

Honkkila, J. and Kavonius, I.K. (2013), “Reconciling micro and macro data on household wealth: a test based on three euro area countries”, *Journal of Economic and Social Policy*, Vol. 15, Issue 2, Article 3.

Horvitz, D.G. and Thompson, D.J. (1952), “A generalization of sampling without replacement from a finite universe”, *Journal of American Statistical Association*, Vol. 47, No 260, pp. 663-685.

Household Finance and Consumption Network, “Questionnaire of the Household Finance and Consumption Survey, Wave 2”, ECB.

Household Finance and Consumption Network, “HFCS core and HFCS non-core variables”, ECB.

Household Finance and Consumption Network (2008a), “Household Finance and Consumption Survey: modalities for implementation”, ECB.

Household Finance and Consumption Network (2008b), “Imputation and data editing”, *HFCN Implementation Documents*.

Household Finance and Consumption Network (2009), “Oversampling the wealthy: eye for an eye, euro for a euro”, *HFCN Implementation Documents*.

Household Finance and Consumption Network (2009), "Survey data on household finance and consumption research summary and policy use", *Occasional Paper Series* No 100, ECB.

Household Finance and Consumption Network (2016): "The Household Finance and Consumption Survey: results from the second wave", *ECB Statistical Paper Series* No. 18.

Jäckle, A., Roberts, C. and Lynn, P. (2006), "Telephone versus face-to-face interviewing: mode effects on data quality and likely causes", *ISER Working Paper,* Series 2006-41.

Kennickell, A.B. (1991), "Imputation of the 1989 Survey of Consumer Finances: stochastic relaxation and multiple imputation".

Kennickell, A.B. (1998), "Multiple imputation in the Survey of Consumer Finances".

Kennickell, A.B. (2005), "The good shepherd: sample design and control for wealth measurement in the Survey of Consumer Finances", presented at the January 2005 *Luxembourg Wealth Study Conference*, Perugia, Italy.

Kennickell, A.B. (2006), "How do we know if we aren't looking? An investigation of data quality in the 2004 SCF", paper prepared for the *2006 Annual Meetings of the American Statistical Association*, Seattle, Washington.

Kennickell, A.B. (2007), "The role of over-sampling of the wealthy in the Survey of Consumer Finances", *Survey of Consumer Finances Working Papers*, July, Federal Reserve Board.

Kennickell, A.B. and Lane, J. (2007), "Measuring the impact of data protection techniques on data utility evidence from the Survey of Consumer Finances", *FRB Working Paper*.

Lehtonen, R. and Pahkinen, E. (2004), "Practical methods for design and analysis of complex surveys", 2nd edition, *Statistics in Practice*, John Wiley and Sons.

Little, R.J.A. and Rubin, D. (2002), "Statistical analysis with missing data", 2nd edition, Wiley, New York.

Little, R.J.A. (1991), "Inference with survey weights", *Journal of Official Statistics*, Vol. 7, No 4, pp. 405-424.

Lohmann, H. (2011), "Comparability of EU-SILC survey and register data: the relationship among employment, earnings and poverty", *Journal of European Social Policy*, Vol. 21, No 1, pp. 37-54.

Mach, L., Saïdi, A. and Pettapiece, R. (2007), "Study of the properties of the Rao-Wu bootstrap variance estimator: what happens when assumptions do not hold?", *Proceedings of the Survey Methods Section*, *SSC Annual Meeting*, June 2007.

OECD (2013): OECD Guidelines for Micro Statistics on Household Wealth, OECD Publishing.

Osiewicz, M. and Pérez-Duarte, S. (2012), "Flexible and homogenous variance estimation in a cross-country survey under confidentiality constraints", Q2012 *European Conference on Quality in Official Statistics*, Athens, May 2012.

Pérez-Duarte, S., Sánchez Muñoz, C. and Törmälehto, V.-M. (2010), "Re-weighting to reduce unit nonresponse bias in household wealth surveys: a cross-country comparative perspective illustrated by a case study", paper prepared for the *Conference on European Quality in Statistics*, Helsinki, 4-6 May 2010.

Preston, J. (2009), "Rescaled bootstrap for stratified multistage sampling", *Survey Methodology*, Vol. 35, No 2, pp. 227-234.

Rao, J.N.K, and Wu, C.F.J. (1988), "Resampling inference with complex survey data", *Journal of the American Statistical Association*, Vol. 83, pp. 231-241.

Rao, J.N.K, Wu, C.F.J. and Yue, K. (1992), "Some recent work on resampling methods for complex surveys", *Survey Methodology*, Vol. 18, No 2, pp. 209-217.

Rao, J.N.K. (1996), "On variance estimation with imputed survey data", *Journal of the American Statistical Association*, Vol. 91, No. 434.

Reiter, J.P., Raghunathan, T.E. and Kinney, S.K. (2006), "The importance of modeling the sampling design in multiple imputation for missing data", *Survey Methodology*, Vol. 27, pp. 85-96.

Royston, P. (2004), "Multiple imputation of missing values", *Stata Journal*, Vol. 4, pp. 227-241.

Rubin, D.B. (1987), "Multiple imputation for nonresponse in surveys", John Wiley & Sons inc.

Rubin, D.B. (1996), "Multiple imputation after 18+ years", *Journal of the American Statistical Association*, Vol. 91, No 434 (Jun.), pp. 473-489.

Sanchez Muñoz, C. (2011), "The Euro-area Household Finance and Consumption Survey – survey mode, oversampling wealthy households and other methods to reduce non-response bias", presented at the *UNECE Conference of European Statisticians*, 2011.

Särndal, C.-E. (2007), "The Calibration Approach in Survey Theory and Practice", *Survey Methodology*, Vol. 33, No 2, pp. 99-119.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992), "Model assisted survey sampling", *Springer Series in Statistics*.

Skinner, C.J. and Holmes, D.J. (1998), "Estimating the re-identification risk per record in microdata", *Journal of Official Statistics*, Vol. 14, Issue 4, pp. 361-372.

Skinner, C.J. and Shlomo, N. (2008) “Assessing identification risk survey microdata using log-linear models”, *Journal of the American Statistical Association*, Vol. 103, Issue 483.

Slacalek, J. (2009), “What drives personal consumption? The role of housing and financial wealth”, *The B.E. Journal of Macroeconomics*, Vol. 9, Issue 1, p. 37, Berkeley Electronic Press.

UN (2001), “Handbook on Population and Housing Census Editing”, *Studies in Methods*, Series F, No 82, United Nations Statistics Division.

UN (2005), “Designing household survey samples: practical guidelines”, *Studies in Methods*, Series F, No 98, United Nations Statistics Division.

UN (2008), “Principles and recommendations for population and housing censuses (Revision 2)”, *Statistical papers*, Series M, No 67, United Nations Statistics Division.

UNECE (2011), “Canberra Group Handbook on Household Income Statistics”, Second Edition, United Nations.

United Nations (2005), “Designing household survey samples: practical guidelines”, *Studies in Methods*, Series F, No 98.

Verma, V. and Betti, G. (2008), “Data accuracy in EU-SILC”, in Atkinson, A.B. and Marlier, E. (ed.), *Income and Living Conditions in Europe*, Eurostat.

Vermeulen, P. (2014), “How fat is the top tail of the wealth distribution?”, *European Central Bank Working Paper Series*, No 1692, July 2014.

Zhang, J.L., Rubin, D.B. and Mealli, F. (2009), “Likelihood-based analysis of causal effects of job-training programs using principal stratification”, *Journal of the American Statistical Association*, Vol. 104, Issue 485, pp. 166-176.

Abbreviations

Countries

AT	Austria	IT	Italy
BE	Belgium	JP	Japan
BG	Bulgaria	LT	Lithuania
CH	Switzerland	LU	Luxembourg
CY	Cyprus	LV	Latvia
CZ	Czech Republic	MT	Malta
DK	Denmark	NL	Netherlands
DE	Germany	PL	Poland
EE	Estonia	PT	Portugal
IE	Ireland	RO	Romania
ES	Spain	SE	Sweden
FI	Finland	SI	Slovenia
FR	France	SK	Slovakia
GR	Greece	UK	United Kingdom
HR	Croatia	US	United States
HU	Hungary		

Others

CAPI	Computer Assisted Personal Interview	ISCED	International Standard Classification of Education
CATI	Computer Assisted Telephone Interview	ISCO	International Standard Classification of Occupations
CAWI	Computer Assisted Web Interview	MI	multiple imputation
ESA	European System of Accounts	NACE	European Classification of Economic Activities
EU-SILC	EU Statistics on Income and Living Conditions	NCB	national central bank
FKP	Financially knowledgeable person	NSI	national statistical institute
HFCN	Household Finance and Consumption Network	PSU	primary sampling unit
HFCS	Household Finance and Consumption Survey	RP	reference person
HMR	household main residence	UDB	User Database

Keywords: survey; wealth; assets; liabilities; households; distribution

JEL codes: C83, D31

© European Central Bank, 2016

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from www.ecb.europa.eu and from RePEc: Research Papers in Economics. Information on all of the papers published in the ECB Statistics Paper Series can be found on the ECB's website.

ISSN 2314-9248 (pdf) DOI 10.2866/260900 (pdf)
ISBN 978-92-899-2643-0 (pdf) EU catalogue No QB-BF-16-011-EN-N (pdf)