EUROPEAN CENTRAL BANK

EUROSYSTEM

STATISTICS PAPER SERIES

NO 1 / APRIL 2013

# THE EUROSYSTEM HOUSEHOLD FINANCE AND CONSUMPTION SURVEY

# METHODOLOGICAL REPORT FOR THE FIRST WAVE


In 2013 all ECB publications feature a motif taken from the €5 banknote.

EUROSYSTEM HOUSEHOLD FINANCE AND CONSUMPTION NETWORK

**Eurosystem Household Finance and Consumption Network**
This report has been prepared by the members of the Eurosystem Household Finance and Consumption Network (see the Annex for the list of members). You can reach us at: hfcs@ecb.europa.eu



**Address** Kaiserstrasse 29, 60311 Frankfurt am Main, Germany
**Postal address** Postfach 16 03 19, 60066 Frankfurt am Main, Germany
**Telephone** +49 69 1344 0
**Internet** http://www.ecb.europa.eu



**ISSN** 2314-9248 (online)
**EU Catalogue No** QB-BF-13-001-EN-N (online)



Information on all of the papers published in the ECB Statistics Paper Series can be found on the ECB's website: http://www.ecb.europa.eu/pub/scientific/stats/html/index.en.htm

# CONTENTS

# ABSTRACT

This report summarises the methodologies used in the first wave of the Eurosystem Household Finance and Consumption Survey, which provides household-level data collected in a harmonised way in 15 euro area countries for a sample of more than 62,000 households. The report presents the methodologies applied in areas such as data collection, sample design, weighting, imputation, and variance estimation. It also analyses issues like differential unit and item non-response and other issues that may have an effect on the comparability of the survey data across countries.



## Keywords

Household-level data, wealth, survey methodology, variance estimation, multiple imputation, statistical disclosure control

## JEL-codes

D12, D14, D31

# I INTRODUCTION

*In 2008, the Governing Council of the European Central Bank decided to conduct a household finance and consumption survey (HFCS) in the euro area. The HFCS provides the Eurosystem with micro-level data on euro area households' wealth and consumption expenditure.*

The Eurosystem Household Finance and Consumption Survey (HFCS),[1] a joint project of all of the central banks of the Eurosystem, provides detailed household-level data on various aspects of household balance sheets and related economic and demographic variables, including income, pensions, employment, gifts and measures of consumption.

A key distinguishing feature of the HFCS is that it provides country-representative data, which have been collected in a harmonised way in 15 euro area members for a sample of more than 62,000 households. Consequently, the survey is unique in that it makes it possible to undertake detailed analysis of issues related to wealth in a fashion that allows consistent comparisons across countries.

When working with the HFCS data, whether using the euro area statistics or undertaking cross-country comparisons, it is important to keep in mind that carrying out such a large survey presents significant conceptual and practical challenges. One difficulty, in particular, is that the survey fieldwork could not be carried out at the same period of time in all countries and, thus, wealth (and income) sometimes refers to different years. Some differences also exist, for example, in the sample selection. Additionally, some structural country differences are not captured in the survey, for example concerning the statutory pension systems. As a result, cross-country comparisons should be made with care and sources of differences should be carefully examined. Even with these caveats, the HFCS initiative is distinctive in its focus on providing ex-ante harmonised wealth data from a large number of countries, thus allowing cross-country in-depth analysis of household finances in the euro area.

This document provides a summary of the main methodological features of the survey, with special emphasis on those that affect cross-country comparability. It is intended to provide a solid background to potential users of the HFCS micro dataset. A companion document, "The Eurosystem Household Finance and Consumption Survey – Results from the First Wave", provides complementary information about key stylised facts emerging from the first wave of the HFCS, tabulating and summarising numerous aspects of wealth heterogeneity across socio-demographic and cross-country dimensions.

The document is in 10 chapters. The remainder of Chapter 1 describes why this kind of data is important for central banks and how they routinely use it for research and policy purposes, and provides a general overview of the HFCS's main methodological features, which are then further expanded in the following chapters. Chapter 2 describes on the HFCS blueprint questionnaire. Chapter 3 documents different data collection approaches and fieldwork issues of the HFCS country surveys. Chapter 4 describes the different sample designs of HFCS country surveys. Chapters 5 and 6 deal with unit non-response / weighting and with item non-response,

[1] See the survey web site, http://www.ecb.int/home/html/researcher_hfcn.en.html, for detailed documentation of the HFCS, including a set of additional descriptive statistics, and for access to the data.

respectively, including a description of the multiple imputation used in the survey as well as data editing. Chapter 7 presents the use of replication-based methods in the HFCS for variance estimation. Chapter 8 describes statistical disclosure control measures applied to the HFCS micro data. Chapter 9 summarises the most important comparability issues to be taken into account when working with the HFCS micro data. Finally, the appendices to the report provide supportive documentation on key HFCS definitions (of household or of the main respondent or financially-knowledgeable person), along with information on the country coverage of core and non-core items in the first wave of the HFCS, on the comparisons between the HFCS and other data sources, and on disclosure control.

## 1.1 WHY CENTRAL BANKS CONDUCT A SURVEY OF HOUSEHOLD FINANCES IN THE EURO AREA

The on-going and long-lasting economic crisis has made it more evident than ever that large structural imbalances may remain hidden behind macroeconomic aggregates. On numerous occasions the early detection of such imbalances requires access to highly granular or even micro-level information. In turn, access to distributional information becomes of the essence.

For the household sector, survey data provide information that permits economic and monetary analyses focusing on particular sub-populations of interest such as on wealthy/poor households, high/low-income households, (highly) indebted households, credit-constrained households, etc. For instance, the implications of the steep increases in household indebtedness in most euro area countries prior to the on-going crisis might have been easier to understand if micro-level data had been available. Had information on over-indebtedness by specific categories of households (as well as on the distribution of debt across income and age classes) been available, central banks could have been better able to detect threats to households' financial soundness and resulting risks to household consumption and to the banking sector.

The Household Finance and Consumption Survey (hereinafter, HFCS) provides data on the assets and debts of individual households in euro area countries. These data allow important insights into the economic behaviour of households, and therefore are a valuable input into a number of Eurosystem policy areas, feeding in particular into monetary policy and financial stability analysis.

The analysis of micro data on households' assets, debts, income and consumption helps improve central banks' knowledge about the economy. Information about the distribution of wealth, debt and income is important for understanding the implications of macroeconomic shocks. For instance, a change in interest rates affects differently the consumption levels of savers and borrowers. Hence, the response of aggregate consumption to such shocks ultimately depends on the percentage of households that are indebted, the level of their debt relative to their income, and the type of debt held (Household Finance and Consumption Network, 2009).

Similarly, for households that have assets, the level and the type of assets held as well as the availability of collateral determine how income, interest-rate, or exchange-rate shocks propagate and affect aggregate demand.

Sometimes the behaviour of certain subpopulations may disproportionately determine the evolution of aggregate statistics. For instance, the accumulation of debt by certain groups of households can have important effects on macroeconomic variables. This was at the origin of the US sub-prime crisis. On the other hand, typically only few, very wealthy households hold certain types of sophisticated products (structured assets, financial derivatives, etc.) or actively participate in stock exchange markets. Their financial decisions may sometimes significantly influence the price of certain assets.

Another example relates to the effects of wealth on consumption: prior to the crisis, the steep increases in housing prices in many advanced economies exerted sizeable effects on the capacity of households to consume, either because households could take additional debt, because they felt they had additional resources available to consume, or because their liquidity or collateral constraints were substantially relaxed.

With the start of the crisis, abrupt declines in the price of real estate not only brought about a sudden halt in household consumption in some countries, but also seriously jeopardised the capacity of the most vulnerable households to cope with their financial commitments.

To properly analyse such effects and to anticipate their possible implications, survey data provide useful input. Since precautionary saving or credit constraints differently affect households with different characteristics, it is important to be able to undertake a differentiated analysis to ultimately determine the size of the response of consumption to shocks, including shocks to wealth. By allowing a comparison of consumption responses to wealth changes across groups differently affected by credit access or risk, micro data facilitate inferences about the motives which drive the transmission of shocks into aggregate demand.

Variations in asset prices and aggregate consumption are partially driven by the same factors, and this makes it difficult to grasp cause-effect relationships at the aggregate level. This problem is much less severe in household-level data because variation between households alleviates identification problems, thus making it easier to find appropriate indicators.

Finally, estimations with aggregate data require relatively long historical series, and may therefore be vulnerable to possible instabilities over time in the underlying relationships. Household-level data are crucial for estimating structural relationships, for instance between consumption and wealth.

Analyses of household-level data reveal that wealth effects differ substantially across households. First, wealth fluctuations bring about asymmetric reactions in household consumption: indeed, households' reaction to losses tend to be more pronounced than to gains in wealth.[2] Besides, housing wealth effects prove to be considerably larger than financial wealth effects.[3] All in all, as has been made evident throughout the on-going crisis, the downturn in the housing market has exerted severe consequences on consumption, largely coming from high rates of home ownership accompanied by significant levels of household indebtedness in several euro area countries.

[2] As shown by Engelhardt (1996) for the US and by Berben et al. (2006) for the Netherlands

[3] See Altissimo et al. (2005) and Slacalek (2009).

Similar arguments explain the usefulness of survey data for analysing household debt. Prior to the on-going crisis, financial innovation and deregulation together with a generally stable macroeconomic environment, historically low interest rates and rising house prices in the euro area, resulted in a substantial accumulation of household debt. These developments raised financial stability concerns related to households' exposure to macroeconomic and idiosyncratic shocks, such as increases in interest rates and unemployment, and variability in asset prices. In this environment, the use of household-level survey data is particularly helpful. For instance, loan and borrower characteristics are important determinants of debt sustainability and of the tendency towards delinquency in serving debt. On the other hand, not all households hold debt. Consequently, the analysis of household debt sustainability should be restricted to the sub-population of indebted households and especially to those who find it difficult to service their commitments.[4] Hence, for both reasons, the risk exposure of the household sector cannot be accurately assessed by looking at aggregate household balance sheet data alone.

## 1.2 GENERAL FEATURES OF THE HFCS

### 1.2.1 THE HOUSEHOLD FINANCE AND CONSUMPTION NETWORK

At the end of 2006, the ECB Governing Council set up the Household Finance and Consumption Network (HFCN).[5] The network is composed of researchers, statisticians and survey specialists from the ECB, the Eurosystem 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 comprises developing and conducting the Eurosystem Household Finance and Consumption Survey (HFCS) and acting 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.[6]

### 1.2.2 GENERAL DESCRIPTION OF THE HFCS

The HFCS is being conducted in a decentralised fashion. Each institution participating in the HFCN (NCB or NSI) is responsible for conducting the survey, and 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.[7]

---

4 Rising aggregate debt levels, for instance, could result from an increasing number of indebted households, or from an increase in leverage of already indebted households, with very different consequences for vulnerability.

5 A predecessor group of experts (the Task Force on a Survey on Household Finance and Consumption) prepared the grounds for the work of the HFCN throughout 2006.

6 The first wave of HFCS was conducted in 15 euro area countries; the survey will start being conducted in all euro area member states, including Ireland and Estonia, as of the second wave of the survey.

7 See Box 1.1.

The HFCS is conducted every three years in most countries[8]. The fieldwork for the first wave was carried out in most countries in 2010 and 2011[9], and most countries intend to conduct the second and third waves in 2014 and 2017, respectively. Table 1.1 provides a summary snapshot of the institution responsible for the HFCS in each country and the fieldwork periods.

**Table 1.1 Main features of the HFCS country surveys**

| Country | Responsible institution | Fieldwork period | Frequency (years) |
|---|---|---|---|
| Belgium | National Bank of Belgium | 2010 | Three |
| Germany | Deutsche Bundesbank | 2010/2011 | Two/Three |
| Estonia | Bank of Estonia | n.a. | Three |
| Ireland | Central Bank of Ireland | n.a. | Three |
| Greece | Bank of Greece | 2009 | Three |
| Spain | Banco de España | 2008/2009 | Three |
| France | Insee / Banque de France | 2009/2010 | Three |
| Italy | Banca d'Italia | 2010 | Two |
| Cyprus | Central Bank of Cyprus | 2010 | Three |
| Luxembourg | Banque centrale du Luxembourg | 2010/2011 | Three |
| Malta | Central Bank of Malta | 2010/2011 | Three |
| Netherlands | De Nederlandsche Bank | 2010 | Three |
| Austria | Oesterreichische Nationalbank | 2010/2011 | Three |
| Portugal | INE Portugal / Banco de Portugal | 2010 | Three |
| Slovenia | Banka Slovenije | 2010 | Three |
| Slovakia | Národná banka Slovenska | 2010 | Three |
| Finland | Statistics Finland / Suomen Pankki | 2010 | Three |

### 1.2.3 METHODOLOGICAL FEATURES OF THE HFCS

#### Ex-ante comparability through an output-oriented approach

When compared with other international initiatives on household wealth surveys,[10] one of the most distinctive features of the HFCS is that the country wealth surveys which are part of the project follow an ex-ante harmonised methodology. In particular, all HFCS surveys provide survey variables according to a set of common definitions and descriptive features according to an output-oriented approach.

Conversely, substantial cross-country differences within the euro area 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, financial markets and products in each country.

8 The HFCS is carried out with a two-year frequency only in Italy, while in Germany it may eventually be conducted every two or every three years.

9 Except in France and Finland, where the survey was conducted in 2009 and 2010, in Greece (2009) and in Spain (2008/2009).

10 Such as the Luxembourg Wealth Study.

Nonetheless, a common Eurosystem blueprint questionnaire is the starting point for country questionnaires.

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 throughout a gradual process. In Cyprus and Portugal the wealth surveys existing in each country were discontinued and replaced by the HFCS.

The HFCS is composed of a common set of core output variables, which all countries report to the ECB according to the agreed common standards and definitions. In addition, there is a set of standardised non-core extensions that countries may voluntarily collect, and which therefore also provide comparable output but only for those countries that collect the information. Country surveys can also collect country-specific (i.e. not necessarily comparable) variables, but these are not included in the euro area HFCS dataset.

Where there is a pre-existing survey, the gradual convergence process implies that for the first wave, a few variables result from combination/adaptation of the original survey variables.

## Sample design

Household samples have been designed in each country to ensure both euro area and country representative results.[11,12] 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 62,000 households were surveyed in the first wave, with varying samples sizes across countries (see further details under chapter 4).

All HFCS country surveys have a probabilistic sample design.[13] This means that each household in the target population should have 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.

## 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 the total mass of wealth. For the former target it is optimal that the sample proportionally represents the

[11] Except in SI, where the reduced sample size of the first wave may not be deemed fully representative for the country.

[12] 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.

[13] Quota sampling was applied in the first wave in Slovakia. As of the second HFCS wave, Slovakia will also adopt a probabilistic sampling approach.

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.[14]

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 distribution of wealth.

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, a large number of countries participating in the HFCS oversample the wealthy via different methods, as described in Table 1.3. The effectiveness of the oversampling is further analysed in chapter 4. All in all, oversampling wealthy households increases precision and reduces non-response bias. In addition, it also improves efficiency in the estimation of variables positively correlated with wealth.

**Table 1.2 Oversampling of the wealthy**

| Country | Oversampling wealthy households | Basis for oversampling |
|---|---|---|
| Belgium | Yes | Geographical areas |
| Germany | Yes | Geographical areas |
| Greece | Yes | Geographical areas |
| Spain | Yes | Taxable wealth |
| France | Yes | Net wealth |
| Italy | No | n.a. |
| Cyprus | Yes | Electricity bills |
| Luxembourg | Yes | Labour income, self-employed |
| Malta | No | n.a. |
| Netherlands | No | n.a. |
| Austria | No | n.a. |
| Portugal | Yes | Geographical areas |
| Slovenia | No | n.a. |
| Slovakia | No | n.a. |
| Finland | Yes | High-income employees, self-employed and farmers |

## Panel component

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

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

surveys in the first group already have a longitudinal component, while some of the others have also initiated or plan to set up a panel as of the second HFCS wave.

The specific situation of each country survey is shown in Table 1.4

**Table 1.3 Panel component**

| Country | Currently | Plans for the future |
|---|---|---|
| Belgium | N.a. | Yes |
| Germany | N.a. | Yes |
| Greece | No | No |
| Spain | Yes | Yes |
| France | No | Yes |
| Italy | Yes | Yes |
| Cyprus | N.a. | Not yet decided |
| Luxembourg | N.a. | No |
| Malta | N.a. | Not yet decided |
| Netherlands | Yes | Yes |
| Austria | N.a. | No |
| Portugal | N.a. | Not yet decided |
| Slovenia | N.a. | No |
| Slovakia | N.a. | Yes |
| Finland | No | No |

Notes: In the "Currently" column, "N.a." indicates countries that have not carried out a comparable wealth survey before, "No" indicates countries continuing a previous survey with a new sample, and "Yes" indicates countries that have households that were interviewed in a previous wave of the survey.

## 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 logical or institutional inconsistencies, such as mistyped or erroneous answers (e.g. amounts or frequencies). To that aim, there is an intensive use of the comments and the paradata provided by interviewers at the conclusion of each interview.[15]

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

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

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 or not correctly collected. 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.

### Box 1.1 Validation and coherence checks of the HFCS data

Data from the various countries participating in the HFCS were screened and edited before transmission to the ECB. Before publication of results from the combined data, a thorough set of validation routines centralised at the ECB were performed on the datasets of all 15 countries participating in the first wave of the HFCS. The validation process was implemented in two steps: first, the technical and logical integrity of the datasets was checked. The second part of the validation process (which can only be run once the technical validation is successfully passed) consists of checking the analytical coherence and plausibility of the data sets, largely via comparisons with external data sources.

The first (technical) part of the validation process was split into critical and informative checks. Critical validation involved checking that all variables, including flag variables, exist and have valid values, that the values of data and flag variables are logically coherent, and that the filtering rules of the survey were respected. If either of these conditions was not met, national institutions in charge of data production were requested to correct and re-submit the data. Informative checks (which did not necessarily imply the need for a new data transmission) involved detecting outlier values of individual variables; high-leverage influential observations for main aggregates; and possible inconsistencies between variables; as well as analysing both survey weights and replicate weights. The results of these informative checks were provided to national institutions in charge of data production for further investigation and verification.

The second (analytical coherence and plausibility) part of the validation process included comparisons of survey results with external sources. For demographic information the main benchmarks were population statistics and the EU-Statistics of Income and Living Conditions (SILC). The total number and the distribution of persons by age were compared to both population statistics and EU-SILC. Additional comparisons with EU-SILC were made to analyse coherence on household characteristics (number of households, household size distribution) as well as on some personal level characteristics, such as education and labour status. Possible conceptual differences, for example of the household definitions between the two surveys, were acknowledged during this comparison.

Aggregate survey values of food consumption and employee income were compared with National Accounts, and aggregate values of deposits and mortgages collateralised by the household main residence were compared with figures from available balance sheet statistics for monetary financial institutions. National institutions in charge of data production were asked to investigate and comment on these figures. Differences in the definitions of households and variables between the HFCS and the benchmark data sources were taken into consideration during the coherence checks.

The results of some of these coherence and plausibility checks are provided in Appendix 10.4.4.

Finally, national institutions in charge of data production were asked to compare the results of the main survey aggregates with their own publications and calculations. Especially in cases where the comparisons rendered substantially different results between national and euro area publications, exhaustive explanations about possible differences in the national definitions or calculation methods which could justify such differences were required.

Since this was the first wave of a very complex survey, much work was eventually necessary to adjust technically all national versions of the HFCS to the common euro area HFCS format. The process of checking data validity and subsequent correction of errors was an intensive one involving a great deal of interaction between the ECB and each national institution. On average, checking and correcting each individual transmission took one month. All in all, before reaching the status of a final validated dataset, several new data transmissions of individual country datasets were necessary. From the first to the last data transmission, the whole process lasted six months on average.

### 1.2.4 CONTINUOUS SURVEY EVALUATION, THE NEED FOR FUTURE RESEARCH AND THE VARIANCE-BIAS TRADE-OFF

Although some surveys that have become a 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.

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 important influence on the bias and variance of estimates based on final data.

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

In the case of protocols for survey execution, there are important known dimensions of differences, which are recorded in this methodological report; for example, substitution sampling was used in some countries and the degree of control of the substitution varied. As regards the statistical processing, the HFCS established high-level frameworks and in some instances made fairly detailed prescriptions. But inevitably, there is room for interpretation and judgment, 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 deeply 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. For example, the trimming of weights for outliers typically lowers the measured variance of final estimates, but at the expense of introducing a formal bias relative to the original sample design; such adjustments are sometimes reasonable, but clear criteria – which are described in this report – are needed to avoid distorting the results. There are similar trade-offs in other aspects of statistical processing, including adjustments for unit nonresponse, imputation, variance estimations procedures, and other areas.

It should therefore be taken into consideration that data sets which are based on a data production process in which a lot of 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 process of data production 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.

# 2 THE HFCS BLUEPRINT QUESTIONNAIRE

*The HFCS blueprint questionnaire consists of three differentiated 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 QUESTIONAIRE

### 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 script providing detailed instructions 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.[16]

### 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 as to the financial information of the household (assets, debts, consumption, etc.) should thus (only) refer to the household members identified in this initial step.

[16] Further details on the selection script of the FKP are provided in Annex 10.1

For the definition of household, the HFCS uses a variation of the so-called "housekeeping concept".[17] 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. On the contrary, possible other persons with usual residence in the dwelling but not sharing 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.[18]

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 a reference person chosen among the household members. The 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 reference person are necessarily different.

Additionally, the interview reference person defined at the beginning of the interview (i.e. the person around which the household is drawn) may not coincide with 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 the level of individual persons, one person most 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 the characteristics of this reference person. 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. The reference person for statistical outputs is therefore constructed ex-post, based on all the information that has been 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,

[17] 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 in depth discussion of these two concepts.

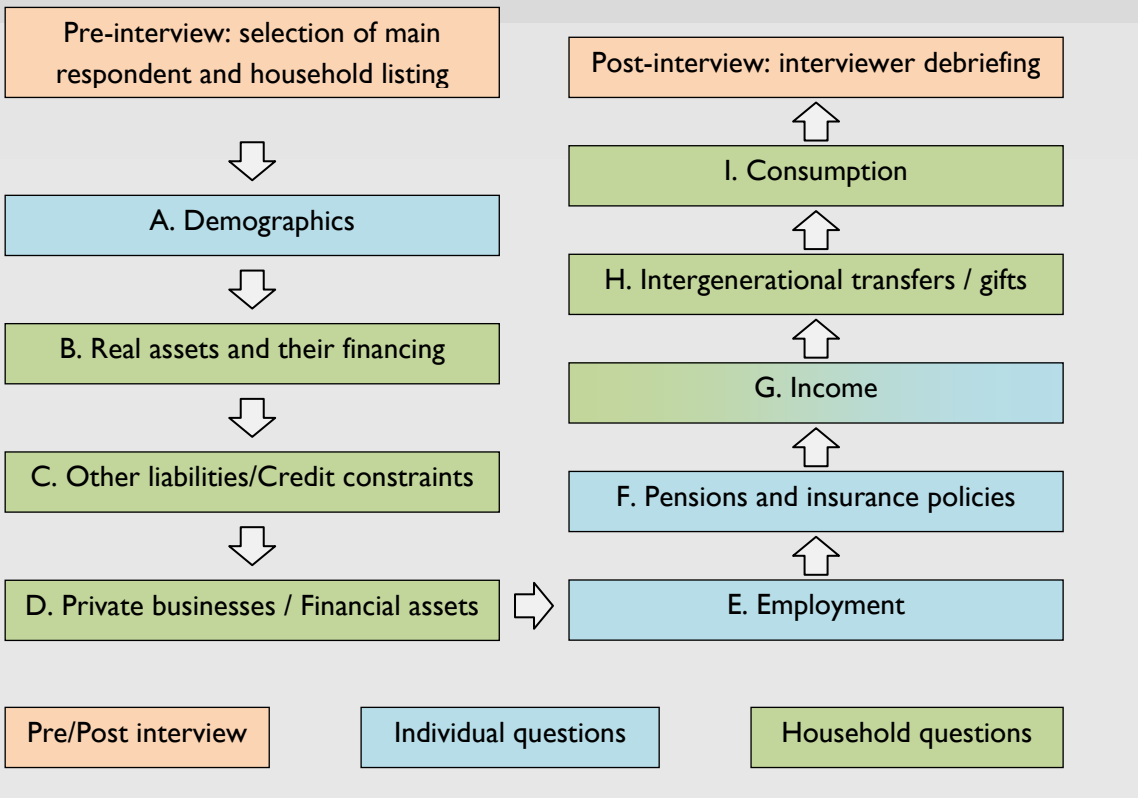
[18] The complete household definition applied for the HFCS is provided in Annex 10.1.

- a lone parent with dependent children,
- the person with the highest income,
- the eldest person.

## 2.2 TOPICS COVERED BY THE HFCS CORE QUESTIONAIRE

The HFCS questionnaire is split into nine sections marked by 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 more. 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).

**Box 2.1 Structure of the HFCS blueprint questionnaire**



### 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+.

### 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 on other real assets. Questions about other characteristics are asked for the household main residence (way and year of acquisition, value at the time of acquisition, etc.). Size of the household main residence and the length of stay in the current household main residence are asked of both owners and tenants. 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), 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.[19]

Selected details containing purpose of the loan, year when the loan was taken 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 and 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

[19] Some of the HFCS countries (Italy, Spain, France) 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.

(applied for credit in the last three years) and credit constraints (credit refusal experience, not applying for credit due to perceived credit constraint).

### 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 ones: sector of activity [NACE[20]], 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, informal loans to relatives or friends, 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) 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 age 16 or over. The first question asks for the self-reported current labour status of each person. A set of questions on main characteristics of employment are asked to persons in employment: status in employment (employee/self-employed/unpaid family worker), occupation (ISCO[21]), 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. A question on previous full- or part-time work is asked to those currently not in employment. Total length of employment is asked to all employed or with previous employment activity.

### 2.2.6 PENSIONS AND LIFE INSURANCE POLICIES

The part on public and occupational pension plans aims at collecting basic information on participation of household members aged 16+ in these types of pension plans, and on the current value of plans with an account balance, if known to the respondent. This particular part of the questionnaire is labelled as indicative, open to particular national implementations.

[20] See http://epp.eurostat.ec.europa.eu/portal/page/portal/nace_rev2/introduction for details of the NACE classification.

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

### 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 main components for the construction of total gross household income, not including lower level details of each of these components (such as, for example, further decomposition 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 Canberra UN expert group of household income statistics.[22] 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. The reason for this is that, given the heterogeneity in taxation and social contributions across countries, the collection of net income figures would endanger cross-country comparability.

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.

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

### 2.2.8 INTERGENERATIONAL TRANSFERS, GIFTS

This section collects information on received inheritances and substantial gifts, aimed to trace household wealth accumulation patterns. The loop for up to the three most important ones 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 on expectations 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 does not intend to collect information on total consumption, but rather on a few consumption indicators that according to the literature[23] may be used to infer total consumption. Collected items include typical monthly amount spent on food at home and typical monthly amount spent on food outside the home over the last 12 months, regular private transfers made outside the household (alimony, assistance, etc.), saving motives, comparison of last 12 moths' expenditure with the

[22] Expert group on Household Income Statistics (2001)

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

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 POSTI INTERVIEW DEBRIEFING/PARADATA

The last part of the questionnaire covers questions intended to close the interview, namely three open questions covering items the respondent found particularly difficult, topics and items that the respondent may have forgotten to report before, and a last one on any other suggestions or comments.

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.

## 2.4 DATA COLLECTION APPROACHES INCORPORATED TO THE QUESTIONNAIRE

### 2.4.1 LOOPS

Loops are sequences of questions referring to individual items, which are repeated for each individual item. There are six loops in the HFCS core questionnaire collecting details on household main residence mortgages, other real estate properties, mortgages on other real estate properties, 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, etc.).[24]

To combat possible respondent fatigue (which could endanger the successful completion of the interview), it is recommended that the CAPI interview tool be equipped with an interactive loop-exit feature. Such an exit feature allows the interviewer to exit the loop and proceed with the mop-up aggregate questions, in this case referring to the items which could not be covered by the detailed questions up to that moment.

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

### 2.4.2 COLLECTION OF MONETARY VALUE QUESTIONS

A standardised CAPI data collection script is used for the collection of monetary values (called the "Euroloop", as it targets the collection of values in euro). The Euroloop encompasses a set of questions which ought to be asked in a strict sequence.

First the interviewer should ask the exact amount, which respondents may provide either in euro or in the 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 in writing to check and confirm.

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

## 2.5 THE HFCS NON-CORE QUESTIONS

The euro area blueprint questionnaire covers the core HFCS survey 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.

By their nature, non-core variables are collected only in a subset of the HFCS countries. Annex 10.3 provides an overview of non-core variables covered in one or more of the HFCS country files in Wave 1.

# 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 a few 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 the collection of 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. For mainly practical or budgetary 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. More specifically, the following are the exceptions to the general rule: while 12 countries applied CAPI interviews in the first wave,

there are three countries in which CAPI was not the main data collection method. In Cyprus, 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 from or estimated with information from administrative registers.

The median duration of the interview was in most countries slightly less than one hour. 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 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 |
|---|---|---|---|---|---|
| Belgium | 100% | 0 | 0 | 0 | n.a. |
| Germany | 100% | 0 | 0% | 0 | 56 |
| Greece | 100% | 0 | 0 | 0 | 43 |
| Spain | 100% | 0 | 0 | 0 | 50 |
| France | 100% | 0 | 0 | 0 | 66 |
| Italy | 85% | 0 | 0 | 15% | 50 |
| Cyprus | 12% | 0 | 0 | 88% | 70 |
| Luxembourg | 100% | 0 | 0 | 0 | 51 |
| Malta | 81% | 0 | 0 | 19% | 39* |
| Netherlands | 0% | 0 | 100% | 0 | n.a. |
| Austria | 100% | 0 | 0 | 0 | 55 |
| Portugal | 100% | 0 | 0 | 0 | 60 |
| Slovenia | 100% | 0 | 0 | 0 | 41 |
| Slovakia | 100% | 0 | 0 | 0 | 52 |
| Finland | 3% | 97% | 0 | 0 | n.a. |

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, interview closure and paradata section.

## 3.2 FIELDWORK

In three countries, Finland, France, and Portugal, the National Statistical Office (NSI) was in charge of data collection and interviews were conducted by staff in the survey units of the corresponding NSIs. 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. In two countries, Cyprus and Germany, the initial survey agency was replaced during the fieldwork period. However, in most cases the experiences with

the survey agencies were positive, and cooperation between NCBs and survey agencies was active and fluent throughout the fieldwork period.

All participating countries organised face-to-face training sessions for interviewers before the start of the fieldwork, with the exception of Netherlands, where no interviewers were used in data collection. In most countries, all interviewers, or at least all new interviewers (in countries adapting the HFCS to an existing survey), participated in the training sessions.

**Table 3.2 Information on interviewer training**

| Country | Length of training, hours | Number of interviewers participating in the training |
|---|---|---|
| Belgium | 6 | 129 |
| Germany | 11 | 230 |
| Greece | 8 | 20 |
| Spain | 28 | 80 |
| France | 27 | 624 |
| Italy | 8 | 192 |
| Cyprus | 5 | 20 |
| Luxembourg | 6 | 41 |
| Malta | 9 | 30 |
| Netherlands | | Not applicable |
| Austria | 7 | 85 |
| Portugal | 16 | 163 |
| Slovenia | 7 | 22 |
| Slovakia | 4 | 50 |
| Finland | 40 | All new interviewers |

Notes: In Finland the training included general interviewer training modules of the NSI.

The number of interviewers varied across countries, to a large extent depending on the sample size. Interviewers were in most cases recruited directly by the survey agency and previous experience in household surveys was a required qualification in all cases. All NCBs or NSIs in charge of data collection participated in the training of interviewers and face-to-face training was organised for a period varying from one day to one week.

Fieldwork periods in the first wave of HFCS varied from three months in Slovakia to eleven months in Germany. Shorter fieldwork periods are beneficial for data comparability, either because the reference periods for income or balance sheet items are closer or, in case of a fixed reference period, to minimise recall bias. On the other hand, longer fieldwork periods allow more opportunities to increase the number of contact attempts and thus to obtain a higher number of interviews. In some cases, initial fieldwork periods had to be extended in order to achieve a sufficient number of interviews.

In ten countries, 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 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 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.

Nine countries (Austria, Belgium, Germany, Greece, France, Italy, Luxembourg, Slovakia, and Spain) conducted pilot surveys prior to the fieldwork in order to test and improve the questionnaire. In Portugal, the training of the interviewers included a test with real interviews with the aim of getting feedback from interviewers on the questionnaire and CAPI program before the fieldwork.

**Table 3.3 Fieldwork indicators**

| Country | Fieldwork company | Number of interviewers conducting the survey | Interviewed households per interviewer | Length of fieldwork period (months) | Adaptation of an existing survey |
|---|---|---|---|---|---|
| Belgium | SA | 129 | 18 | 7 | N |
| Germany | SA | 191 | 19 | 11 | N |
| Greece | SA | 72 | 41 | 4 | N |
| Spain | SA | 80 | 77 | 7 | Y |
| France | NSI | 624 | 24 | 6 | Y |
| Italy | SA | 192 | 41 | 7 | Y |
| Cyprus | SA | 35 | 35 | 9 | N |
| Luxembourg | SA | 38 | 25 | 8 | N |
| Malta | SA | 30 | 28 | 5 | N |
| Netherlands | SA | n.a. | n.a. | 9 | Y |
| Austria | SA | 80 | 30 | 9 | N |
| Portugal | NSI | 125 | 35 | 4 | N |
| Slovenia | SA | 21 | 16 | 3 | N |
| Slovakia | SA | 50 | 40 | 3 | N |
| Finland | NSI | 150 | 73 | 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 HFCS and previous wealth surveys by Statistics Finland.

## 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 and 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 one, 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. Most of these cases refer to the Finnish data. In addition, income variables in France are based on tax files, and 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 3.5. 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 (Törmälehto, Kannas and Säylä 2012). 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. 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. For deposits, there was no reliable register source available, and the values for deposits were constructed with statistical matching, using the 2004 wealth survey as donor data and the 2009 income distribution survey sample as recipient data.

**Table 3.4 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. |
| France | Register data are used to construct income variables. |
| Italy | Legislative and institutional information is used to construct variables on the number of public pension schemes. |
| 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. |
| Finland | Register data: All income variables except private transfers, outstanding amounts of liabilities, loan payments, ownership and number of cars and other vehicles, ownership and values of mutual funds, bonds and quoted shares, and education.<br>Estimated data: Value of household main residence, ownership and value of other properties, values of cars and other vehicles, business wealth, ownership and values of deposits, and values and contributions to voluntary pension schemes. |

# 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.[25] 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.

To broaden the analytical potential of the HFCS data in the future, a panel component is foreseen in several of the participating countries. In some countries that adapted HFCS to existing wealth surveys, a panel component was already in use, while some other national wealth surveys already plan to have a panel component in future waves.

[25] Only one country, Slovakia, did not use probability sampling, but quota sampling. For the second wave, all 17 participating countries plan to use probability sampling.

## 4.2 MAIN COUNTRY FEATURES

While all of the main features of the sample design exist in all first-wave surveys, countries have adopted a variety of approaches. 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 measurement objectives of the survey (UN 2005). Units to be interviewed can be selected in one in or multiple stages. In a multiple stage design, the first stage (or stages) involves a selection of geographical areas, from which individual household are chosen in the final stage.

Table 4.1 describes the sampling designs used in various countries. Seven countries used one-stage stratified sampling, while eight countries had a multi-stage stratified sampling design. The sample size of Slovenia in the first wave is only representative at the euro area level, while in all other countries the sample size was chosen to be representative also at the country level.

**Table 4.1 Sampling designs in HFCS**

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

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

In Italy, the Netherlands and Spain, where collection of the HFCS output variables is adapted to existing wealth surveys a panel component is included. The Finnish data constructed for the sample of the EU-SILC survey have a four-year rotational panel design. However, for now it is foreseen to produce the HFCS data as independent cross-sectional data. In addition, a few countries (Belgium, Germany and Slovakia) plan to introduce a panel component for the forthcoming waves.

Table 4.2 describes the sampling frames and 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 income and labour status in Finland and Luxembourg and personal taxable wealth in Spain.

**Table 4.2 Sampling frames and stratification criteria**

| Country | Sampling frame(s) | Stratification criteria |
|---|---|---|
| Belgium | Telephone register and street register | NUTS I region and average income by neighbourhood of residence |
| Germany | Clusters of addresses from municipalities (NSI); list of street sections, population registers of municipalities | Demographic size, average taxable income of municipalities; additionally wealth-related parameters of street sections for large municipalities |
| Greece | List of municipalities (Census); random routing for SSU | NUTS II region, degree of urbanisation |
| Spain* | Municipal census (list of addresses) supplemented by tax office information; list of addresses | Population of the municipality, taxable wealth |
| France | List of geographical units (ZAE, based on Census); list of dwellings | Region (ZAE), regional population; socio-economic criteria |
| Italy* | List of municipalities; resident lists from municipalities | NUTS II region and population of the municipality |
| Cyprus | Customer register of the Electricity authority of Cyprus | Census districts divided into urban and rural |
| Luxembourg | Addresses of fiscal households from social security register | Individual income, nationality, employment status |
| Malta | Dwelling register of the NSI | Statistical region |
| Netherlands* | Postal addresses | NA |
| Austria | List of enumeration districts; register of post-box addresses | NUTS III region, population of municipality |
| Portugal | List of geographical areas; list of private dwellings, from Census | NUTS II region |
| Slovenia | List of districts from Census; list of persons 16+ from population register | Population of the municipality, with adjustments for expected non-response |
| Slovakia | List of municipalities, households chosen by random walk. | NUTS III region, population of municipality. In each stratum, ten income quotas were prescribed, which interviewers had to fulfil |
| Finland* | Central population register using master sample of 50,000 persons 16+ and members living in the same household-dwelling unit | Socio-economic criteria of the highest income-earner |

* Gross sample includes panel households that have responded to previous waves of the same survey

**Table 4.3 Numbers of strata and primary sampling units selected**

| Country | Number of strata | PSUs selected |
|---|---|---|
| Belgium | 24 | - |
| Germany | 3 | 198 |
| Greece | 41 | 169 |
| Spain | 32 | - |
| France | 22 | 525 |
| Italy | 50 | 359 |
| Cyprus | 8 | - |
| Luxembourg | 20 | - |
| Malta | 68 | - |
| Netherlands | n.a. | - |
| Austria | 193 | 422 |
| Portugal | 7 | 800 |
| Slovenia | 6 | n.a. |
| Slovakia | 40 | n.a. |
| Finland | 26 | - |

Note: Number of strata refers to the first sampling stage only. Primary sampling units selected are shown for countries with multistage sampling designs.

### 4.2.2 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, i.e. non-resident citizens were excluded. In addition, in most national surveys all institutionalised population were 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 of the areas of the Republic of Cyprus not under the effective control of the Government of the Republic of Cyprus. In Greece, small villages were left out of the gross sample because of cost considerations. This led to a decrease of approximately 7% in the population frame in Greece compared to population statistics. 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 4.4 Excluded groups**

| Country | Excluded groups |
|---|---|
| Belgium | Homeless, prisoners |
| Germany | Homeless, all institutionalised population |
| Greece | Homeless, all institutionalised population, smaller villages, comprising about 7% of the total number of households |
| Spain | Homeless, all institutionalised population |
| France | All institutionalised population |
| Italy | Homeless, all institutionalised population |
| Cyprus | Homeless, prisoners, population of the areas of the Republic of Cyprus not under the effective control of the Government of the Republic of Cyprus |
| Luxembourg | Diplomats, non-resident citizens, homeless, international civil servants and in general households where no individual is entitled to be registered in the social security register, all institutionalized population |
| Malta | Diplomats, non-resident citizens, armed forces, homeless, civilians living in military institutions, prisoners |
| Netherlands | Blind people, people who do not speak Dutch, all institutionalised population. |
| Austria | Homeless, all institutionalised population |
| Portugal | All institutionalised population, homeless, people living in military area |
| Slovenia | All institutionalised population, diplomats, homeless, non-citizens, armed forces, civilians living in military area |
| Slovakia | Homeless, all institutionalised population |
| Finland | All institutionalised population |

### 4.2.3 USE OF REPLACEMENTS

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

Replacements were used in six countries. This includes Slovakia, which used quota sampling. Although the rules for using replacements varied, all countries followed the criteria mentioned above to a large extent.

In Slovakia, interviewers selected households in each stratum by random walk to fulfil the prescribed quotas for each income group. In Belgium, replacements for households in the telephone register could be contacted after refusal or after ten unsuccessful contact attempts of the original sample unit, and for households in the street register, after refusal or after four unsuccessful attempts. Replacements in the telephone register were randomly selected from the same stratum, and replacements in the street register, from another dwelling in the same street.

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 stratum (within the same zip code for urban areas) after three unsuccessful contacts, on different days and at different times, determining not-at-home, refusals or ineligibility. Slovenia had an original and a replacement sample with exactly the same primary sampling units. The interviewers tried to get a response from two or three households units from each primary sampling unit of the original sample. If this was not achieved, they were allowed to use additional contacts from the same primary sampling unit of the replacement sample. In Cyprus, replacements were selected from the same stratum as the original sample unit.

### 4.2.4 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 make 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 suffer from the inability to collect data from the top fractions of the distribution. This will have an impact on indicators such as the Gini coefficient and quantile ratios (for example, the ratio of net wealth between the households in the top 20 and bottom 20 per cent of the wealth distribution), which are sensitive to the values of the richest households.

Capturing values of assets from the wealthiest households is even more relevant in the case of some 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 expected to be higher for wealthier households. This is first caused by the special difficulty of establishing contact with wealthy respondents, since they are 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 participation in surveys.[26] 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,[27] 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.

Nine out of fifteen countries used different strategies to oversample wealthy households (table 4.5). In addition, Austria oversampled households in Vienna, and Slovenia households in Ljubljana and Maribor because of higher expected non-response rates in those regions. The strategies varied a lot 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 Finland and Luxembourg, individual-level income and in Cyprus, household-level electricity consumption were used as proxies for wealth. In Belgium and Germany, the proxy for wealth was regional-level income and in Greece, regional real estate prices. Other countries oversampled municipalities in which higher income has been observed and in which non-response rates were expected to be higher.

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

[27] 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 4.5 Oversampling strategies**

| Country | Criteria for oversampling | Details |
|---|---|---|
| Belgium | Average regional income | Neyman allocation, based on the standard deviation of income in stratum and stratum size. |
| Germany | Taxable income of regions | Smaller municipalities (population<100,000) and, in larger municipalities, street sections with high average income (>€92,000) are oversampled. |
| Greece | Regional; real estate prices | The sampling rate for Athens and Thessaloniki is proportional to the real estate prices of each cluster. |
| Spain | Taxable wealth of individuals | Eight wealth strata were defined and were oversampled progressively at higher rates. |
| France | Wealth | For the wealthy sample, four strata have been made: wealthy city dwellers, equity-based wealth, real estate-based wealth, lower wealth. For each primary unit and each stratum, an allocation proportional to main residences is computed. Then, a systematic selection is made within each couple stratum-primary unit. |
| Italy | No oversampling | - |
| Cyprus | Electricity consumption | 61 % of the gross sample was drawn from households within the top 10% according to electricity consumption. |
| Luxembourg | Personal income subject to social contributions | 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. |
| Malta | No oversampling | - |
| Netherlands | No oversampling | - |
| Austria | Regional | Some oversampling in Vienna because of higher expected non-response rate. |
| Portugal | Region | Metropolitan areas of Lisbon and Porto oversampled, 50% of gross sample drawn from these areas. |
| Slovenia | Region | Municipalities of Ljubljana and Maribor were oversampled, as higher non-response rates were expected. |
| Slovakia | No oversampling | To help interviewers fulfil the prescribed income quotas, a list of streets with higher incidence of wealthy households in each stratum was provided. |
| Finland | Individual income and socio-economic status from population register | High-income employees, self-employed and farmers are oversampled, as well as “others” and “no tax”. |

The oversampling strategies have enriched the sample with a higher proportion of cases 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 4.6). 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 five and the richest ten percent of the population.

To compute this indicator, the net wealth values of the 90$^{th}$ and 95$^{th}$ percentiles were first calculated from the weighted data. Subsequently, the 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.

The interpretation of the figures in table 4.6 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 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 first wave of the survey. 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 almost 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.

**Table 4.6 Effective oversampling rates of the wealthy**

| Country | Effective oversampling rate of the top 10% | Effective oversampling rate of the top 5% |
|---|---|---|
| Belgium | 47 | 60 |
| Germany | 117 | 148 |
| Greece | -2 | 3 |
| Spain | 192 | 314 |
| France | 129 | 208 |
| Italy | 4 | 0 |
| Cyprus | 55 | 81 |
| Luxembourg | 55 | 66 |
| Malta | -5 | -13 |
| Netherlands | 87 | 98 |
| Austria | 1 | 4 |
| Portugal | 16 | 20 |
| Slovenia | 22 | 31 |
| Slovakia | -11 | -8 |
| Finland | 68 | 85 |

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.

# 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 first 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. Due to the sensitivity of wealth data, observed unit non-response rates have been generally higher in wealth surveys than in income surveys.[28]

To make the survey as comparable as possible, a compatible approach should be used in all countries executing the HFCS. To improve the quality of the analysis to be done 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.

## 5.2 UNIT NON-RESPONSE IN THE HFCS

The sensitivity of the core wealth information in the HFCS may lead some households to be less likely to cooperate in the survey, and cultural differences across countries may cause the degree of such sensitivity to vary. Even households that are potentially interested in participating may sometimes be hard to reach, because they live in a home that is not easily accessible to an

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

interviewer, and such accessibility may differ across wealth groups. Many other factors have the potential for influencing the degree of response and the composition of the set of respondents.

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 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 first wave of the HFCS. These indicators are based on standard definitions proposed on the topic (see AAPOR 2011). The following indicators are included:

1) Response rate = Achieved interviews / Eligible sample units[29]
2) Refusal rate = Sample units refusing to participate / Eligible sample units
3) Cooperation rate = Achieved interviews / Contacted sample units
4) Contact rate = Sample units contacted / Eligible sample units
5) Eligibility rate = Eligible units / Gross sample size

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.

In sample designs where the units are sampled with unequal probabilities, weighted indicators on response behaviour will differ from unweighted ones. Weighted response rates are calculated by weighting the sample units with their inverse selection probabilities (design weights) before the calculation of the indicators. The weighted response rate estimates the percentage of units in the frame that are represented. While weighted response rates measure survey output quality, unweighted response rates measure fieldwork process quality. For the HFCS, unequal probability sampling is most common, given that oversampling of wealthy households is widely used. Weighted response rates thus improve the comparison of survey quality across various countries, since oversampling specific population groups is beneficial for survey quality. But as in the case of oversampling of hard-to-interview wealthy households, oversampling may lead to diminished response rates.

[29] 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.

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 Portugal and especially in Malta. In the comparison of response rates it is noteworthy to mention 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 Spain, Italy, the Netherlands and Finland, the survey also has a panel component. These issues had a positive impact on the response rate. Response rates of households interviewed for the first time are given in the footnote of table 5.1.

**Table 5.1 Response behaviour indicators in the HFCS**

| Country | Gross sample size | Net sample size | Response rate | Weighted response rate | Refusal rate | Cooperation rate | Contact rate | Eligibility rate |
|---|---|---|---|---|---|---|---|---|
| Belgium | 11,376 | 2,364 | 21.8% | n.a. | 57.6% | 27.2% | 80.1% | 95.4% |
| Germany | 20,501 | 3,565 | 18.7% | n.a. | 69.7% | 21.1% | 94.2% | 92.9% |
| Greece | 6,354 | 2,971 | 47.2% | 48.7% | 46.4% | 47.8% | 98.7% | 99.1% |
| Spain | 11,782 | 6,197 | 56.7%* | n.a. | 34.8% | 58.4% | 97.2% | 92.6% |
| France | 24,289 | 15,006 | 69.0% | 69.6% | 30.0% | 69.0% | 100% | 89.8% |
| Italy | 15,592 | 7,951 | 52.1%* | 53.2% | 37.8% | 57.8% | 90.2% | 97.8% |
| Cyprus | 3,938 | 1,237 | 31.4% | 32.4% | 56.6% | 35.7% | 88.0% | 100% |
| Luxembourg | 5,000 | 950 | 20.0% | 19.3% | 63.7% | 21.0% | 95.5% | 94.9% |
| Malta | 3,000 | 843 | 29.9% | 30.4% | 34.1% | 44.3% | 67.5% | 94.0% |
| Netherlands | 2,263 | 1,301 | 57.5%* | n.a. | 42.5% | 57.5% | 100% | 100% |
| Austria | 4,436 | 2,380 | 55.7% | 56.4% | 39.6% | 56.7% | 98.1% | 96.3% |
| Portugal | 8,000 | 4,404 | 64.1% | 59.0% | 10.3% | 80.2% | 79.9% | 85.9% |
| Slovenia | 965 | 343 | 36.4% | 35.6% | 45.9% | 41.6% | 87.5% | 97.8% |
| Slovakia | 2,000 | 2,057 | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. |
| Finland | 13,525 | 10,989 | 82.2%* | 85.0% | 11.1% | 86.2% | 95.4% | 98.7% |

# In France and Portugal, survey participation is compulsory for households.

* Response rates for the whole sample; more comparable response rates are the response rates for households interviewed for the first time, which are 40.3% in Spain, 35.0% in Italy and 70.1% in Finland. This figure is not available for the Netherlands.

## 5.3 WEIGHTING

Weighting procedures are an essential tool for adjusting, to the degree 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 take 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) (ECB, 2011). The methodology is coherent with existing international standards (Eurostat, 2011 and United Nations, 2005). These steps are implemented sequentially as follows:

i. 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.

ii. 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.

iii. The coverage-adjusted weights are further adjusted in an attempt to minimize 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 level of groups.

iv. 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 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 in most countries in the HFCS.

There are some exceptions to this approach, mainly as a consequence of the use of relatively complex sample frames. In Belgium, which employed a dual-frame sampling approach involving both telephone and address registers, different calculation methods were necessary for the design and final weights to account for the higher selection probability of households in the telephone register. In Spain, which conducts its survey as a panel with refreshment samples to maintain cross-sectional representation, the panel and non-panel components of the sample are considered as two independent samples. For panel households, the initial weights were calculated as the design weights from the previous wave (2005) corrected for non-response in

2005. Non-response adjustments were made for both panel and non-panel households within the cells defined by the various sampling frame variables. Before calibration, the two sample components were combined and their weights corrected according to the relative size of the sub-samples.

In Finland the HFCS weights are based on the weights of the income distribution statistics, with some weights having been downscaled due to outlier values detected for some wealth items. The data of the income distribution survey consist of four panels, and design weights are constructed, adjusted for non-response and calibrated separately for each panel. Finally weights are rescaled by the relative sample sizes of each panel.

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 variables (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 efficiency (low variance) of estimates (Little, 1991). In case of highly variable weights the efficiency of the estimates can be increased by the trimming of extreme weights.

Extreme weights were only trimmed in the surveys carried out in Greece and Finland. In the calibration, limits for weight adjustment factors can be set in order to define a ceiling for the ratio between (coverage/non-response adjusted) design weights and final weights. This procedure was applied in Germany, Finland, Luxembourg, Malta, Slovakia, Slovenia and Spain.

### 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 | | | NSI |
| Germany | x | x | x | x | Municipality size, homeownership, size of main residence (for homeowners); education, labour status and nationality (of main income earner) | Micro census |
| Greece | x | | x | x | Homeownership, education | LFS, EU-SILC |
| Spain | x | x | x | | Municipality size | Census 2008 |
| France | x | x | | x | Municipality size, education, type of household, job qualifications, labour and wealth income | NSI |
| Italy | x | x | | x | Municipality size | NSI |
| Cyprus | x | | x | x | - | Census |
| Luxembourg | x | x | x | | Nationality, labour status, stratum | Social security register |
| Malta | x | x | x | x | Employment status | LFS |
| Netherlands | x | x | | | Household population, homeownership, education | EU-SILC, CBS Statline |
| Austria | | | x | | Municipality size | Micro census |
| Portugal | x | x | x | x | Total number of households | Population statistics, LFS |
| Slovenia | x | x | x | | - | Population register |
| Slovakia | x | x | x | x | Homeownership | NSI |
| Finland | x | x | x | x | Selected income variables | Population and tax register |

LFS: Labour force survey. NSI: National statistical institute. EU-SILC: Statistics on Income and Living Conditions. CBS: Central Bureau of Statistics, Netherlands.

* Cell-based poststratification

### 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 | 4,692,601 | 2,017 | 90.3 |
| Germany | 39,673,000 | 11,128 | 99.1 |
| Greece | 4,114,150 | 1,385 | 54.6 |
| Spain | 17,017,706 | 2,746 | 135.3 |
| France | 27,860,408 | 1,857 | 100.3 |
| Italy | 23,818,293 | 2,996 | 103.4 |
| Cyprus | 303,242 | 245 | 80.2 |
| Luxembourg | 186,440 | 196 | 64.5 |
| Malta | 143,677 | 170 | 51.0 |
| Netherlands | 7,386,144 | 5,677 | 93.2 |
| Austria | 3,773,956 | 1,586 | 52.6 |
| Portugal | 3,932,010 | 893 | 83.3 |
| Slovenia | 777,777 | 2,268 | 99.0 |
| Slovakia | 1,911,664 | 929 | 90.1 |
| Finland | 2,531,500 | 230 | 84.4 |

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.

## 5.4 EVALUATION AND FUTURE RESEARCH

As mentioned in the first chapter of this report, the HFCS continuous process of survey evaluation is intended to building up a body of knowledge on survey methodology aspects, such as non-response and weighting. This is necessary in order to understand more deeply the effects of non-response on survey bias, the factors across countries that may influence non-response and non-response bias, and the steps that can reasonably be taken to further minimise the effects of non-response on the survey results used for policy and research.

Because non-response bias may affect the measured distribution of various estimates differently, a broad-spectrum approach is required for investigation of non-response bias. It may also be important to further clarify differences in the structure weighting adjustments across countries, to address the data available for such adjustments, and to investigate any differences in estimated variance across countries that are substantially different from what would be expected based on only the sample design and the response rate.

# 6 ITEM NON-RESPONSE, MULTIPLE IMPUTATION AND EDITING

*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.*

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

## 6.1 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.

### 6.1.1 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 wave, a minimum set of variables that need to be imputed has been determined (Household Finance and Consumption Network, 2008b).[30] This set of 130 variables includes basically all components of household income, consumption and wealth.

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

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

general, the data constructor's responsibility, since "in general, ultimate users have neither the knowledge nor the tools to address missing data problems satisfactorily." Data-base constructors using individual HFCS country data have better information on the reasons for nonresponse 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.

### 6.1.2 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.[31]

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 data set. 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 nonresponse.

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 data sets 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.

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

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).

### 6.1.3 METHODOLOGY AND COMMON SOFTWARE TOOLS

In multinational surveys, countries should use similar methodologies for imputing 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.

A common SAS software package called €MIR has been developed for the purpose of multiply imputing HFCS data. This set of programs includes checks for the logical flow of the questionnaire, 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 program FRITZ 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 some 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), there should be little differences to be expected in the outcome of same imputation models in comparison to €MIR.

A Stata software package called MeDaMi was developed in the network. MeDaMi checks data consistency with the HFCS benchmark, prepares the data for multiple imputation, imputes the data and evaluates imputed data. The imputation routine in MeDaMi is exactly the same (ICE) as in other countries using Stata for imputation. The most substantial contribution of MeDaMi is the automation of the specification of suitable imputation model. For each variable to be imputed the program selects a list of the most significant covariates (significance thresholds are determined by the user) and checks the estimability of the proposed models. 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 incentives of the data producer to fully examine the relationships between different variables, missingness patterns, etc. that are vital in the construction of good quality imputation models.

Of the 15 countries participating in the first wave of the HFCS, 13 used MI to correct for item non-response. The exceptions were Finland and Italy, where 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.[32] 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 with register-based estimation, and the share of missing information for any variable that was collected was negligible.

Descriptions of some of the most important methodological choices for the imputation models are presented in tables 6.1 – 6.2. In table 6.1, the first column shows whether survey weights are used in the imputation models – either by performing weighted regressions or by using survey weights as covariates. The use of survey weights in the imputation models is useful to reduce the bias caused by unequal probabilities of selection in the sample. On the other hand, weighted regression potentially leads to less efficient estimates (Faiella, 2010).

The second issue in table 6.1 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).

Finally, the treatment of outliers – both imputed and observed variables used as covariates – and imputed values for which inconsistencies between variables were detected, are shown. The last item considers observations edited after the imputation process is finished, i.e. editing data values produced by the imputation model. During the imputation process, both outliers and inconsistencies are thoroughly checked in most countries, and the imputation models are modified, if they produce implausible results. Additionally, bounds are introduced in the imputation models of various countries to force the imputed observations to be within

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

reasonable values (for example, to be positive for most kinds of balance sheet items). However, in only few countries, a negligible number of imputed values were exactly equal to the lower or upper bound introduced in the model.

**Table 6.1 Imputation methodology**

| Country | Use of weights | | | Limiting imputation due to low number of observations | | | | Outliers | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Weighted regression | Weight as covariate | No weights used | Yes, values left missing | Yes, estimation/other methodologies used | Yes, variables pooled and imputed | No | Editing of imputed outliers | Editing of inconsistent imputed values | Excluding outliers from covariates | None |
| Belgium | | x | | | | x | | x | x | x | |
| Germany | x | x | | | | | x | | x | x | |
| Greece | | | x | | x | | | x | x | x | |
| Spain | | | x | x | | x | | | | | x |
| France | | | x | | | | x | x | | | |
| Italy | x | | | x | | | | | | x | x |
| Cyprus | | | x | | x | | | x | x | | |
| Luxembourg | | x | | | | x | | | | | x |
| Malta | | | x | | x | | | | | | x |
| Netherlands | | | x | | x | | | | | | x |
| Austria | x | | | x | x | x | | | x | | |
| Portugal | x | | | | x | | | x | x | x | |
| Slovenia | | | x | | x | | | x | x | | |
| Slovakia | | | x | | x | | | x | x | | |
| Finland | Not applicable | | | | | | | | | | |

LFS: Labour force survey. NSI: National statistical institute. EU-SILC: Statistics on Income and Living Conditions. CBS: Central Bureau of Statistics, Netherlands.

* Cell-based poststratification

Table 6.2 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 the degree to which the broad conditioning approach (see chapter 6.1.2) was applied in various countries. 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 collecting this data, numerous dummy variables are often created from regional variables to be used as covariates in the imputation models.

**Table 6.2 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 | 46 | 31 | 49 | 50 |
| Germany | 84 | 10 | 17 | 20 |
| Greece | 233 | 154 | 49 | 196 |
| Spain | 239 | 104 | 159 | 224 |
| France | 17 | 12 | 21 | 7 |
| Italy | n.a. | n.a. | 10 | n.a |
| Cyprus | 50 | 38 | 48 | 98 |
| Luxembourg | 86 | 118 | 31 | 40 |
| Malta | 4 | 10 | 14 | 8 |
| Netherlands | 6 | 7 | 7 | 5 |
| Austria | 104 | 51 | 133 | 102 |
| Portugal | 16 | 23 | 17 | 6 |
| Slovenia | 47 | 4 | 14 | 130 |
| Slovakia | 102 | 31 | 69 | 100 |
| Finland | n.a. | n.a. | n.a. | n.a |

## 6.2 COMPARATIVE INFORMATION ON ITEM NON-REPONSE AND IMPUTATION

This section presents data on the outcome of the imputation process for all 15 countries that participated in the first wave of the HFCS. The first subsection looks at the level of item non-response for the most important variables and the second subsection on the impact of imputation on the aggregate variables. These indicators reflect the degree and quality of imputations in different countries.

### 6.2.1 ITEM NON-RESPONSE RATES FOR MAIN VARIABLES

Tables 6.3 – 6.6 show information on the imputed observations for 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. 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.[33]

[33] As had 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.

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 Luxembourg, Malta and Slovakia, high share of balance sheet and income values have been imputed from a range value provided by the respondent. This procedure should be distinguished from an imputation from 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 staying below 10% in all countries. Imputation rates for employee income also remained quite low in most cases. 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 expected, given the low share of imputed values. In a few cases, the imputed values of savings accounts have a significantly higher mean compared with the collected values. This should indicate that households that were not able to record the values of their savings accounts are expected to possess higher savings accounts values than average households, given the covariates used in the imputation model.

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.

### 6.2.2 IMPACT OF IMPUTATION ON MAIN VARIABLES

For the calculation of aggregate variables, both the share of imputation-affected observations and the share of imputed values in total are of significance. Table 6.7 shows these shares for three aggregate variables: gross real and financial wealth and gross income. The share of observations affected by imputation is the share of households (or of persons over 15 years old for individual variables) for which at least one of the components entering the computation of the aggregate was imputed. Since the importance of various components with respect to the aggregates varies significantly, the second column for each aggregate shows the imputed

value's share of the total. This is the weighted sum of all components of the aggregate that were imputed divided by the weighted sum of the aggregate variable.

Imputation has a bigger impact on financial wealth than on real wealth or income. Households had difficulty providing values for several financial assets, and in seven countries the share of imputed values in total gross financial wealth is more than 20 %. The share of imputed values in real wealth estimates is relatively large (above 20%) in two countries (Cyprus and Austria), but for most other countries less than 10 %. Also, the share of imputed values in gross income estimates is less than 10 % in the majority of countries.

**Table 6.3 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 | 74.0 | 0.1 | 92.4 | 4.6 | 2.4 | 273,100 | 272,800 |
| Germany | 56.0 | 0.4 | 89.9 | 5.3 | 3.6 | 205,800 | 206,200 |
| Greece | 66.8 | 0.0 | 91.0 | 5.6 | 3.3 | 123,400 | 124,100 |
| Spain | 86.9 | 0.0 | 90.9 | 4.9 | 4.3 | 211,100 | 212,300 |
| France | 66.7 | 0.0 | 0.0 | 80.9 | 19.1 | 222,200 | 230,200 |
| Italy | 71.2 | 0.0 | 99.4 | 0.0 | 0.0 | | - |
| Cyprus | 80.0 | 0.0 | 81.8 | 0.0 | 17.5 | 317,500 | 334,700 |
| Luxembourg | 70.0 | 0.0 | 88.3 | 8.9 | 2.9 | 611,900 | 611,500 |
| Malta | 76.3 | 0.0 | 67.5 | 30.3 | 0.2 | | - |
| Netherlands | 74.1 | 0.0 | 94.5 | 0.0 | 5.5 | 270,600 | 269,900 |
| Austria | 48.4 | 1.2 | 74.8 | 15.3 | 9.1 | 258,100 | 258,600 |
| Portugal | 69.4 | 0.0 | 90.0 | 0.0 | 7.0 | 113,800 | 115,700 |
| Slovenia | 82.2 | 0.9 | 82.1 | 7.4 | 10.5 | 126,500 | 128,600 |
| Slovakia | 77.3 | 0.0 | 81.1 | 14.6 | 4.1 | 68,700 | 69,200 |
| Finland | 77.0 | 0.0 | All values estimated | | | | - |

* 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 edited, estimated or collected as range values and then imputed. Provided only for countries with >15 imputed cases.

**Table 6.4 Item non-response rates: Highest 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 | 27.9 | 0.2 | 78.0 | 8.7 | 13.0 | 79,100 | 81,800 |
| Germany | 22.0 | 0.7 | 89.8 | 3.7 | 4.6 | 73,400 | 73,900 |
| Greece | 13.5 | 0.0 | 77.6 | 0.0 | 22.4 | 50,700 | 50,900 |
| Spain | 18.7 | 0.1 | 91.1 | 4.4 | 4.5 | 72,700 | 73,200 |
| France | 14.6 | 0.0 | 75.4 | 0.0 | 24.5 | 73,200 | 75,400 |
| Italy | 7.9 | 0.0 | 99.5 | 0.0 | 0.0 | | - |
| Cyprus | 44.2 | 0.1 | 87.6 | 0.0 | 10.4 | 105,000 | 98,400 |
| Luxembourg | 34.0 | 0.5 | 85.7 | 6.4 | 7.0 | 157,900 | 160,500 |
| Malta | 10.7 | 0.0 | 67.8 | 22.2 | 2.2 | | - |
| Netherlands | 53.3 | 1.2 | 88.9 | 0.0 | 11.1 | 125,500 | 123,200 |
| Austria | 15.0 | 1.0 | 59.8 | 19.9 | 20.3 | 66,500 | 56,400 |
| Portugal | 23.4 | 0.0 | 79.6 | 0.0 | 20.2 | 54,600 | 54,900 |
| Slovenia | 8.7 | 0.5 | 62.9 | 0.0 | 37.1 | | |
| Slovakia | 11.1 | 0.0 | 67.7 | 17.5 | 14.8 | 26,600 | 27,200 |
| Finland | 37.0 | 0.0 | 100.0 | 0.0 | 0.0 | | - |

* 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 edited, estimated or collected as range values and then imputed. Provided only for countries with >15 imputed cases.

**Table 6.5 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 | 79.1 | 0.6 | 79.8 | 9.2 | 10.7 | 48,600 | 40,800 |
| Germany | 81.2 | 0.3 | 86.3 | 6.0 | 7.8 | 22,500 | 22,800 |
| Greece | 3.6 | 0.1 | 78.1 | 5.5 | 16.4 | 44,400 | 35,800 |
| Spain | 30.9 | 0.1 | 85.4 | 5.3 | 9.3 | 37,000 | 37,200 |
| France | 87.3 | 0.0 | 30.9 | 63.5 | 5.6 | 15,200 | 15,400 |
| Italy | 25.6 | 0.0 | 49.2 | 50.8 | 0.0 | | - |
| Cyprus | 36.5 | 0.2 | 90.0 | 0.0 | 10.0 | 50,800 | 51,000 |
| Luxembourg | 74.1 | 0.2 | 57.9 | 24.8 | 16.6 | 39,700 | 42,400 |
| Malta | 78.4 | 5.3 | 33.7 | 46.0 | 20.1 | 27,400 | 23,700 |
| Netherlands | 85.3 | 1.6 | 96.9 | 0.0 | 3.1 | 23,500 | 23,700 |
| Austria | 85.7 | 1.6 | 63.6 | 18.3 | 18.0 | 30,100 | 29,000 |
| Portugal | 42.6 | 0.2 | 87.1 | 0.0 | 12.9 | 28,200 | 25,200 |
| Slovenia | 29.7 | 0.3 | 69.9 | 11.7 | 18.4 | 14,300 | 12,200 |
| Slovakia | 26.4 | 0.2 | 62.6 | 22.8 | 14.6 | 7,800 | 7,800 |
| Finland | 0.0 | 100.0# | All values estimated | | | | - |

* 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 edited, estimated or collected as range values and then imputed. Provided only for countries with >15 imputed cases.

**Table 6.6 Item non-response rates: Gross employee income**

| 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 | 47.2 | 0.4 | 76.2 | 6.5 | 7.5 | 32,100 | 31,200 |
| Germany | 44.5 | 3.5 | 32.2 | 2.8 | 11.1 | 27,600 | 27,600 |
| Greece | 36.1 | 0.1 | 92.7 | 1.8 | 5.4 | 18,300 | 18,300 |
| Spain | 36.2 | 0.2 | 91.4 | 4.4 | 4.2 | 17,400 | 17,500 |
| France | 42.7 | 0.0 | 100.0 | 0.0 | 0.0 | | - |
| Italy | 32.8 | 0.0 | 100.0 | 0.0 | 0.0 | | - |
| Cyprus | 50.6 | 0.2 | 90.5 | 0.0 | 9.5 | 24,600 | 24,600 |
| Luxembourg | 52.5 | 0.5 | 70.3 | 20.0 | 8.4 | 48,900 | 48,900 |
| Malta | 42.8 | 0.0 | 52.6 | 43.2 | 2.7 | 14,800 | 14,900 |
| Netherlands | 41.3 | 3.3 | 76.4 | 8.6 | 14.9 | 31,900 | 31,500 |
| Austria | 44.4 | 5.1 | 71.3 | 11.5 | 15.9 | 25,600 | 24,300 |
| Portugal | 38.9 | 0.8 | 94.2 | 0.0 | 5.4 | 11,900 | 11,800 |
| Slovenia | 46.7 | 0.3 | 75.8 | 13.0 | 11.2 | 14,100 | 14,900 |
| Slovakia | 62.3 | 0.2 | 73.5 | 18.8 | 7.7 | 6,300 | 6,400 |
| Finland | 63.7 | 0.0 | 100.0 | 0.0 | 0.0 | | - |

* 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%. If net income values were reported instead of gross income, the figures refer to net income.

# Includes observations edited, estimated or collected as range values and then imputed. Provided only for countries with >15 imputed cases.

**Table 6.7 Impact of imputation on aggregate variables**

| Country | Gross real wealth | | Gross financial wealth | | Gross income | |
|---|---|---|---|---|---|---|
| | Share of imputation-affected observations | Share of imputed observations in total | Share of imputation-affected observations | Share of imputed observations in total | Share of imputation-affected observations | Share of imputed observations in total |
| Belgium | 11.4 | 5.5 | 26.0 | 31.2 | 13.9 | 12.7 |
| Germany | 10.8 | 6.6 | 13.5 | 16.8 | 24.4 | 11.7 |
| Greece | 9.0 | 4.8 | 16.5 | 25.1 | 8.7 | 6.6 |
| Spain | 17.2 | 6.2 | 16.6 | 10.9 | 13.0 | 3.6 |
| France | 18.9 | 13.5 | 9.7 | 4.7 | 0.0 | 0.0 |
| Italy | 0.8 | 0.2 | 9.9 | 4.5 | 0.0 | 0.0 |
| Cyprus | 44.5 | 24.5 | 49.6 | 28.3 | 19.3 | 12.6 |
| Luxembourg | 13.0 | 5.5 | 31.1 | 16.7 | 19.6 | 8.8 |
| Malta | 4.7 | 10.5 | 36.0 | 24.4 | 39.6 | 6.2 |
| Netherlands | 17.0 | 11.1 | 43.0 | 43.4 | 28.1 | 19.9 |
| Austria | 13.7 | 28.6 | 27.2 | 29.8 | 34.6 | 20.0 |
| Portugal | 11.0 | 8.0 | 16.4 | 20.5 | 10.7 | 7.3 |
| Slovenia | 17.2 | 16.2 | 20.1 | 18.2 | 21.1 | 12.9 |
| Slovakia | 10.2 | 5.3 | 17.4 | 20.9 | 11.1 | 6.8 |
| Finland | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

Share of imputation-affected observations: share of households for which at least one component of the aggregate variable was imputed.
Share of imputed observations in total: weighted sum of all components of the aggregate that were imputed divided by the weighted sum of the aggregate variable.

## 6.3 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 coefficient 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 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 6.8).

**Table 6.8 Information on data editing**

| Country | Organisation responsible for editing | Interviewer comments used in editing | Re-contacting of households possible* | Number of households re-contacted |
|---|---|---|---|---|
| Belgium | NCB and SA | Systematically | Yes | n.a. |
| Germany | NCB | Systematically | No | - |
| Greece | NCB | In most cases | Yes | n.a. |
| Spain | NCB | Systematically | Yes | 800 |
| France | NCB and NSI | Systematically | No | - |
| Italy | NCB and SA | Systematically | Yes | n.a. |
| Cyprus | NCB | In most cases | Yes | n.a. |
| Luxembourg | NCB and SA | Systematically | No | - |
| Malta | NCB | Occasionally | Yes | n.a. |
| Netherlands | SA | Not applicable | Yes | n.a. |
| Austria | NCB and SA | Systematically | Yes | 400 |
| Portugal | NCB | Systematically | Yes | n.a. |
| Slovenia | NCB and SA | Occasionally | Yes | 13 |
| Slovakia | NCB | No | No | - |
| Finland | NSI | Systematically | Yes | n.a. |

* Only re-contacts for verification of data values included, re-contacting households for verification of data authenticity excluded.
Notes: NCB; National Central Bank, NIS: National Statistical Institute, SA: Survey Agency

Table 6.9 shows the shares of edited observations for the variables included in tables 6.2 – 6.5, and the number of variables with relatively high edit rates. For most of the cases there was limited editing for the main variables. However, in two countries over 10% of observations for employee income included some editing.

**Table 6.9 Edit rates**

| Country | Value of main residence | Highest mortgage on main residence | Savings accounts | Employee income | Number of variables with edit rates >5% |
|---|---|---|---|---|---|
| Belgium | 1% | 0% | 0% | 10% | 28 |
| Germany | 1% | 2% | 0% | 11% | 37 |
| Greece | 0% | 0% | 0% | 0% | 5 |
| Spain | 0% | 0% | 0% | 0% | 0 |
| France | 0% | 0% | 0% | 0% | 0 |
| Italy | 0% | 0% | 0% | 0% | 1 |
| Cyprus | 1% | 2% | 0% | 0% | 11 |
| Luxembourg | 0% | 1% | 1% | 1% | 12 |
| Malta | 2% | 8% | 0% | 2% | 25 |
| Netherlands | 0% | 0% | 0% | 0% | 2 |
| Austria | 1% | 0% | 0% | 1% | 12 |
| Portugal | 0% | 0% | 0% | 0% | 0 |
| Slovenia | 0% | 0% | 0% | 0% | 27 |
| Slovakia | 0% | 0% | 0% | 0% | 0 |
| Finland | 0% | 0% | 0% | 0% | 0 |

# 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 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 has been addressed in the chapter 6 on Imputation, and which will be connected to total variance estimation in this chapter.*

*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 7.4.*

## 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.

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 both with different kinds of sampling designs and with 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[34] sample of PSUs from each stratum is selected.[35] 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 generalized 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 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 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 $h = 1,...,H$, 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 $y_{hi}$

[34] 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.

[35] 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.

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:

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

2. 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 $\lambda_h = \sqrt{\frac{m_h(1-f_h)}{n_h - 1}}$.

3. 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 $V_*(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}_{*b} - \bar{\hat{\theta}}_*)^2$, 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 (HW0010) for each observation is then rescaled by a specific factor $n_h / (n_h - 1)$, 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 practise 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.

**Table 7.1 Variance estimation model**

| Country | Institution responsible for replicate weights | Stages | Strata | Certainty units |
|---|---|---|---|---|
| Belgium | NCB | Sampling design | Sampling design | No |
| Germany | NCB | 1st stage | 1st stage | No |
| Greece | NCB | 1st stage | 1st stage | No |
| Spain | NCB | 1st stage | 1st stage | No |
| France | NSI | 1st stage | 1st stage | Yes |
| Italy | NCB | 1st stage | 1st stage | Yes |
| Cyprus | NCB | Sampling design | Sampling design | No |
| Luxembourg | SA | Sampling design | Sampling design | No |
| Malta | NCB | Sampling design | Sampling design | No |
| Netherlands | SA | Sampling design | Sampling design | No |
| Austria | NCB | 1st stage | 1st stage | No |
| Portugal | NSI | 1st stage | 1st stage | No |
| Slovenia | NCB | 1st stage | 1st stage | No |
| Slovakia | NCB | Sampling design | Sampling design | No |
| Finland | NSI | Sampling design | Sampling design | No |

Notes: Stages and Strata refer to the sampling stages that were used in the calculation of the replicate weights. If the replicate weights use the same stages and strata, it is indicated with "sampling design"; otherwise the stratification criteria and the stages used are indicated.

### 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. In all countries, the replicate weights all sum up to the same number of households. In most countries, they sum up to the same number of persons, depending on the exact calibration used, and when they do not, the variation of the number of persons is limited.

**Table 7.2 Calibration of replicate weights**

| Country | At household level | At person level | By gender | By age group |
|---|---|---|---|---|
| Belgium | Yes | Yes | Yes | Yes |
| Germany | Yes | No (0.2%) | No (0.9%) | No (1.4%) |
| Greece | Yes | No (0.1%) | No (0.9%) | No (1.6%) |
| Spain | Yes | No (1.1%) | No (1.4%) | No (2.9%) |
| France | Yes | No (0.3%) | No (3.2%) | No (0.6%) |
| Italy | Yes | No (0.6%) | No (0.7%) | No (0.9%) |
| Cyprus | Yes | No (5.4%) | No (4.3%) | No (1.9%) |
| Luxembourg | Yes | No (0.2%) | No (0.9%) | No (1.5%) |
| Malta | Yes | Yes | Yes | No (2.0%) |
| Netherlands | Yes | Yes | No (1.7%) | No (1.7%) |
| Austria | Yes | No (1.1%) | No (1.6%) | No (3.4%) |
| Portugal | Yes | Yes | Yes | Yes |
| Slovenia | Yes | Yes | Yes | No (4.4%) |
| Slovakia | Yes | Yes | Yes | Yes* (0.8%) |
| Finland | Yes | Yes | Yes | Yes |

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.
*: different age groupings were used 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 Perez-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 7 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 $\theta_m$ 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 $\theta^*_{mb}$, with mean across replicates $\bar{\theta}_m = \frac{1}{B}\sum_{b=1}^{B}\theta^*_{mb}$. The partial variance for implicate $m$ is $U_m = \frac{1}{B-1}\sum_{b=1}^{B}\left(\theta_{mb} - \bar{\theta}_m\right)^2$. 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, $Q = \frac{1}{M-1}\sum_{m=1}^{M}\left(\theta_m - \bar{\theta}\right)^2$ 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 $\left(\theta - \bar{\theta}\right)T^{-1/2}$ is approximately distributed as a $t$-distribution with $\upsilon_M$ degrees of freedom, with $\upsilon_M = (M-1)\left(1 + \frac{W}{(1+1/M)Q}\right)^2$.

Barnard and Rubin (1999) recommend an alternative measure in the case of small samples, since in that case the $\upsilon_M$ can be much larger than the complete data degrees of freedom. This

recommended measure is $\upsilon_M^* = \left( \frac{1}{\upsilon_M} + \frac{1}{\upsilon_{obs}} \right)^{-1}$, where $\upsilon_{obs} = \frac{\upsilon_0 + 1}{\upsilon_0 + 3} \upsilon_0 (1 - \gamma)$, $\upsilon_0$ is the complete-data degrees of freedom, and $\gamma = \frac{(1 + 1/M)Q}{T}$.

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

### 7.4.1 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. However, at least up to version 11.2, there is no official 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.

A modified Stata command, which replaces an internal routine used by **mi estimate**, was run before the estimation command.[36] The command suppresses an internal check in the Stata command which forbids users from running **mi estimate** while **mi svyset** is set to use replicate weights. The modified procedure produces the correct standard errors, according to the methodology outlined above.

In Stata 12, the use of an undocumented option "**vceok**" may allow the standard routine to proceed. It is used as: **mi estimate, vceok: svy ...**.

[36] This modified Stata routine is available on the website of the Household Finance and Consumption network, http://www.ecb.int/home/html/researcher_hfcn.en.html.

**Table 7.3 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: svy: mean da1110
```

### 7.4.2 APPLICATION IN SAS

The SAS statistical system has several routines starting in 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 7.4 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 needs thus to be renamed.

```
proc surveyreg data=HFCS varmethod=brr(fay=0.000);
    model da1110 = da1120;    * model;
    repweights wr0001-wr1000; * replicate weights;
    by _Imputation_;          * 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, the comparability of the data. Country-specific anonymisation techniques may also be centrally applied by the ECB in close coordination with the NCB (NSI) concerned, where necessary to ensure the confidentiality of the replies.

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 foreseen 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.[37]

It consists of the following techniques:

- The following variables are kept unchanged: household identification number, country and type of dwelling. In the case of a panel survey, the following variables are kept unchanged: past household ID, vintage of last interview and survey vintage. If the household identification number has not been randomised by the Member State, it (and its panel equivalent) is randomised. 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.

[37] 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 10.4.

#### Demographics

Age is top-coded at 85 years. In a few countries, only age in 5-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..

#### Financial assets

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 some additional variables have been top-coded or recoded into coarser categories in order to reduce identity disclosure risk.

### 8.1.3 EFFECT OF TOP-CODING ON ESTIMATES

The following table shows the percentage of applicable cases that reach the top-coding value, for selected variables.

**Table 8.1 Cases in the top-coded range – Categorical variables (%)**

| Number of… | other properties | cars | motorbikes | trucks | vans |
|---|---|---|---|---|---|
| Belgium | 0.6 | 3.4 | 1.4 | 2.7 | 0.9 |
| Luxembourg | 5.8 | 4.3 | | | |
| Malta | 4.5 | 5.2 | | | |
| Portugal | 4.1 | 1.6 | 0.6 | 1.9 | 1.1 |

Notes: Percentages are shown for all households holding the asset. Countries not shown did not apply top-coding of these categorical variables; empty cells mean that the variable was not top-coded in the corresponding country.

**Table 8.2 Cases in the top- or bottom-coded range age-related variables (%)**

| Country | Age RA0300 | Length in residence HB0200 | Year of acquisition HB0700 | Inheritances received HH0201 |
|---|---|---|---|---|
| Belgium | 1.76 | 0.13 | 0.00 | 1.62 |
| Germany | 1.52 | 0.79 | 0.05 | 0.00 |
| Greece | 1.10 | 0.07 | 0.05 | 0.00 |
| Spain | 2.96 | 0.07 | 0.07 | 0.00 |
| France | 1.91 | 0.23 | 0.01 | 0.01 |
| Italy | 2.41 | 0.21 | 0.12 | n.a. |
| Cyprus | 0.60 | 0.24 | 0.10 | 0.00 |
| Luxembourg | 0.59 | 0.00 | 0.30 | 0.00 |
| Malta | 1.34 | 0.47 | 0.16 | 0.40 |
| Netherlands | 0.30 | 0.00 | 0.00 | 0.00 |
| Austria | 1.79 | 0.34 | 0.00 | 0.00 |
| Portugal | 1.93 | 0.09 | 0.00 | 0.22 |
| Slovenia | 1.14 | 0.29 | 0.00 | 0.00 |
| Slovakia | 0.22 | 0.00 | 0.00 | 0.00 |
| Finland | 0.87 | n.a. | n.a | n.a. |

Notes: Percentage of cases in each country in the top- or bottom-code.

## 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 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 roughly 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 can only be applied when there is a clear case of reidentification 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 minimal also on medians.

**Table 8.3 Rounding of variables in nominal amounts**

| Data range (USD in the SCF, EUR in the HFCS) | SCF rounding to the nearest… | Rounding to 2 digits, to the nearest… | Rounding to 3 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.

## 8.4 ADDITIONAL COUNTRY-SPECIFIC PROCEDURES

Some countries have applied additional anonymisation procedures. In particular, RA0200 (gender) has been deleted in LU for household members under 15 years old.

In Spain the country of birth (RA0400) has been dropped for confidentiality reasons.

# 9 COMPARABILITY ISSUES

*One of the goals of the HFCS project is to ensure as much as possible that the euro area 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.*[38]

## 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 THE 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).

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 the different countries ranges from November 2008 to August 2011. The reference periods for assets and liabilities match within each country, and range from the end of 2008 to the beginning of 2011. The reference periods for income cover 2007, 2008, 2009, or the 12 months before the interview.

[38] 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.

Table 9.1 Reference periods

| Country | Fieldwork | Assets & Liabilities | Income |
|---|---|---|---|
| Belgium | 04/10 – 10/10 | Time of interview | 2009 |
| Germany | 09/10 – 07/11 | Time of interview | 2009 |
| Greece | 6/09 – 9/09 | Time of interview | Last 12 months |
| Spain | 11/08 – 07/09 | Time of interview | 2007 |
| France | 10/09 – 02/10 | Time of interview | 2009 |
| Italy | 01/11 – 08/11 | 31/12/2010 | 2010 |
| Cyprus | 04/10 – 01/11 | Time of interview | 2009 |
| Luxembourg | 09/10 – 04/11 | Time of interview | 2009 |
| Malta | 10/10 – 02/11 | Time of interview | Last 12 months |
| Netherlands | 04/10 – 12/10 | 31/12/2009 | 2009 |
| Austria | 09/10 – 05/11 | Time of interview | 2009 |
| Portugal | 04/10 – 07/10 | Time of interview | 2009 |
| Slovenia | 10/10 – 12/10 | Time of interview | 2009 |
| Slovakia | 09/10 – 10/10 | Time of interview | Last 12 months |
| Finland | 01/10 – 05/10 | 31/12/2009 | 2009 |

The time dimension has an effect on comparability, since the amounts are nominal and do not include inflation. One possibility is to correct for these differences with the inflation rate. The most straightforward approach would use the Harmonised Index of Consumer Prices to calculate differences. Given the low inflation in the euro area in the period of the different country surveys (with a maximum correction of 4.7% in one country), taking inflation into account does not change the overall picture by much (see table 9.2).

Nevertheless, 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 first 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.

### 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 Possible Inflation and Purchasing power parity correction factors

| Country | Date of reference | HICP | PPP |
|---|---|---|---|
| Belgium | 2010 | 1 | 0.850 |
| Germany | 2010 | 1 | 0.907 |
| Greece | 2009 | 1.047 | 1.041 |
| Spain | 2009 | 1.024 | 1.025 |
| France | 2009 | 1.0174 | 0.842 |
| Italy | 2010 | 1 | 0.914 |
| Cyprus | 2010 | 1 | 1.072 |
| Luxembourg | 2010 | 1 | 0.796 |
| Malta | 2010 | 1 | 1.305 |
| Netherlands | 2009 | 1.0093 | 0.875 |
| Austria | 2010 | 1 | 0.862 |
| Portugal | 2010 | 1 | 1.150 |
| Slovenia | 2010 | 1 | 1.145 |
| Slovakia | 2010 | 1 | 1.413 |
| Finland | 2009 | 1.0169 | 0.801 |

Notes: HICP: Harmonised Index of Consumer Prices, Overall index, calculated to adjust values in the HFCS to 2010 amounts.
Source: Eurostat (2010) for HICP and purchasing power parity factors, HFCN calculations.
How to read: euro amounts in Belgium should be multiplied by 0.850 and by 1.305 in Malta 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 two countries, collection is by other means (telephone and web), with approximately 20% of the interviews in two other countries with face-to-face paper questionnaires. Finally, in one country there is a predominance of paper questionnaires.

In two countries, a significant share of information is provided through registers, which might make comparisons difficult (see e.g. Lohmann, 2011, for an analysis of the case of register data in EU-SILC). One particular difficulty is that registers might not match exactly the common survey definition; for example, in the case of Finland, self-employment businesses are proxied by the ownership (and the value) of unquoted shares.

### 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.

In no country is the common questionnaire completely implemented. The table below lists the number of variables in the core questionnaire that are available (including all three loops). A total of 383 variables in the household-level file and 58 in the person-level file were foreseen, although this is an overestimate of the number of "useful" variables, since some variables are deducted from others, or were applicable only to a very limited number of households or persons.

**Table 9.3 Variables Available in the User DataBase (UDB)**

| Country | Household-level file | Personal-level file |
|---|---|---|
| Belgium | 298 | 55 |
| Germany | 375 | 56 |
| Greece | 382 | 53 |
| Spain | 373 | 53 |
| France | 313 | 54 |
| Italy | 261 | 45 |
| Cyprus | 383 | 58 |
| Luxembourg | 296 | 56 |
| Malta | 308 | 52 |
| Netherlands | 288 | 46 |
| Austria | 383 | 52 |
| Portugal | 379 | 58 |
| Slovenia | 301 | 56 |
| Slovakia | 380 | 58 |
| Finland | 40 | 32 |

Notes: number of variables having at least one non-empty observation.

Most countries provided most of the variables in both the household- and personal-level files. In Finland, due to the nature of the data collection stemming primarily from registers, many detailed variables are not available (e.g. description of assets, separate values of items collected in loops), though aggregates are provided (see section 3.3).

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-88 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 nine 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. In

Greece, this approach was used for the collection of employee income. Respondents had the option to provide net income for all income components in Slovenia and for some income components in Austria (see table 9.4), in which case gross income was estimated. 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. France had access to tax income registers and was able to provide taxable gross income.[39]

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

**Table 9.4 Deviations in the collection of income variables**

| Country | Information |
|---|---|
| Belgium | Legal information was used to transform net income into gross if the interviewer indicated that households could not report gross income. |
| Germany | Gross income collected, but respondents had the option to provide net income for employee, self-employment, capital and pension income, and unemployment benefits. If provided, net income figures where converted to gross income using information from the tax system. |
| France, Finland | Income data derived from registers. |
| Greece | Employee income collected in net amounts, gross employee income constructed by adding estimated taxes and social security contributions. Other income variables collected in gross terms. |
| Italy | Income collected net of taxes and social contributions. Gross income estimated from net income. |
| Austria | If respondents were not able to provide gross amounts, net income was collected for employee, self-employment and pension income, These net income data were transformed to gross income using information of net income, employment status, household structure (taking care of deductibles for child provisions) and geographical location of household. |
| Slovenia | Gross income collected, but respondents had the option to provide net income. In the latter case, gross income was estimated. |

### 9.2.6 EDITING

As described in chapter 6, editing aims at manually correcting the cases where the information has been erroneously recorded.

The major cause 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, household structure. In the absence of sufficiently detailed information, which would be prohibitively expensive to collect in a face-to-

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

face survey, the conversion requires a number of assumptions, which might limit comparability not only across countries, but also across households inside each country.

The estimation of financial income, i.e. income earned from financial assets, was carried out in one country, given that collecting this information directly from households is often met with poor 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.[40] Table 9.5 lists the number of variables in the HFCS core variables in the to-be-imputed list that contain more than 10 observations whose value should have been imputed but was not.

**Table 9.5 Number of Variables in the to-be-imputed list with missing or not collected values**

| Country | Date of reference | | HICP PPP | |
|---|---|---|---|---|
| | not fully collected | not fully imputed | not fully collected | not fully imputed |
| Belgium | 0 | 0 | 0 | 0 |
| Germany | 5 | 12 | 0 | 4 |
| Greece | 0 | 0 | 0 | 0 |
| Spain | 7 | 0 | 2 | 0 |
| France | 19 | 24 | 0 | 2 |
| Italy | 17 | 1 | 0 | 0 |
| Cyprus | 0 | 0 | 3 | 0 |
| Luxembourg | 0 | 0 | 0 | 0 |
| Malta | 0 | 0 | 0 | 0 |
| Netherlands | 3 | 8 | 0 | 0 |
| Austria | 0 | 0 | 0 | 0 |
| Portugal | 0 | 0 | 0 | 0 |
| Slovenia | 0 | 0 | 0 | 0 |
| Slovakia | 0 | 0 | 0 | 14* |
| Finland | 69 | 1 | 0 | 1 |

Notes: variables not fully collected count the number of variables having at least some observations not collected, and do not include items collected in the third loop. Third loop items were not collected in several countries, though the correct remainder item has been provided. Multiple secondary labour statuses are also excluded from the totals. Variables not fully imputed include variables for which more than ten observations have not been imputed.

* Employment variables were not collected for persons aged 16 in Slovakia.

[40] 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 given to provide researchers with a set of least common variables, for example, in the case of age (coarsened to 5-year brackets in some countries), by providing the coarsened variable for all countries.

# 10 APPENDICES

## 10.1 HFCS DEFINITIONS OF FINANCIALLY KNOWLEDGEABLE PERSON AND HFCS HOUSEHOLD DEFINITION

### 10.1.1 HFCS DEFINITION OF FINANCIALLY KNOWLEDGEABLE PERSON (FKP)

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.

### 10.1.2 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 the essentials of living. Employees of other residents (i.e. live-in domestic servants, au-pairs, etc.) and roommates 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 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 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 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.[41]

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 month' 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 accommodations, then whether they are members of this (or their other) household depends on the length of their absence.

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. Particularly 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.

[41] 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.

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 institutionalised population are considered out of 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.

## 10.2 COVERAGE OF THE CORE ITEMS IN THE FIRST WAVE OF THE HFCS

While the surveys of countries starting new surveys or replacing their previous survey with the new HFCS survey to supply the HFCS data are largely built on the euro area blueprint questionnaire, several countries (Italy, Spain, Finland, France, the Netherlands) have adapted existing national surveys to the HFCS. In the countries with pre-existing surveys, the harmonisation may involve some approximation until the survey gradually converges to the output variables of the Eurosystem HFCS survey. The whole content of the HFCS blueprint questionnaire may therefore not always be fully covered in all countries as of Wave1 of the survey.

The following table provides information on the incomplete coverage of the HFCS core questions in the first wave of the HFCS.

**Table 10.1 HFCS core variables not covered in HFCS Wave1**

**DEMOGRAPHICS**

Questions on country of birth and length of stay in the country for the foreign born are not collected in France or the Netherlands. In Cyprus, the items on gender, marital status and education were only collected for the interview reference person.

**REAL ASSETS AND THEIR FINANCING**

The question on how long the household has lived in the current household main residence is not asked in Finland. The question on the amount of rent paid for partially owned household main residence is not asked in France, Italy or the Netherlands. The question on % ownership of household main residence is not asked in Finland or the Netherlands. The question on the way household main residence was acquired is not asked in Finland and is only partially asked in France. The question on the year of household main residence acquisition is not asked in Finland. The question on household main residence value at the time of its acquisition is not asked in France or Finland and is only partially asked in Italy.

The questions on vehicles are not asked in France. The questions on the number of cars and the number vehicles other than cars are not asked in Italy – the current value of all vehicles is provided jointly in the value item for cars, which includes cars as well as other vehicles together. The question on the number of vehicles other than cars is only partially asked in Spain. The questions on the ownership and value of valuables (such as jewellery, works of art, antiques) are not asked in Finland.

The questions on individual household main residence mortgages are not asked in Finland, one total register-based outstanding amount of household main residence mortgages and monthly payment on mortgages is provided for all mortgages together. The questions on other property mortgages are not asked separately in Finland, and are included together with non-mortgage loans in the 'Other liabilities' section. 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) – type of guaranty.

The purpose of the household main residence and other property mortgages is only partially asked in Spain and France. The questions on additional borrowings made on mortgages are not asked in Italy or the Netherlands, and are only partially asked in Spain. The question on mortgage refinancing is not asked in Italy and is only partially asked in Spain.

**OTHER LIABILITIES, CREDIT CONSTRAINTS**

The questions on leasing are not asked in Italy. The questions on credit card debt are not asked in France. The questions on leasing, credit lines/overdrafts and credit card debt are not asked in Finland. The questions on non-collateralised loans are not asked in Finland; one total register-

based outstanding amount is provided for all non-collateralised loans together. Only one main purpose of the non-collateralised loans is asked in Italy; secondary purposes are not collected. The purpose of the loan question for non-collateralised loans is only partly asked in Spain. The amount initially borrowed and initial length of the loan for non-collateralised loans are not collected in Italy. The current interest on non-collateralised loans is only partly collected in Italy.

The questions on credit applications and credit constraints are not asked in Finland. The questions on credit applications and credit constraints are not provided for Italy due to the different approach and wordings used in the national questionnaire, which are not directly comparable with the HFCS output. The questions on credit refusal and on non-application for credit due to perceived credit constraint are only partly asked in Spain. The question on re-application for credit after refusal is not asked in Spain.

**PRIVATE BUSINESSES, FINANCIAL ASSETS**

The questions on self-employment businesses are not asked in Finland; one total register-based value of self-employment businesses is provided for all self-employment businesses together. The questions on ownership and value of non-self-employment not publicly traded businesses are not asked in Finland.

The question on sector of activity of self-employment businesses (NACE) is not asked in Italy. The question on household members working in the self-employment businesses is only partly asked in Italy.

The value of saving accounts is not separately collected in Finland and is provided jointly with the sight accounts. Sub-items of the mutual funds questions for mutual funds predominantly investing in real estate and hedge funds are not separately collected in Italy. The question on types of owned bonds is not asked in Finland, France or Italy. The question on foreign shares in the owned shares' portfolio is not asked in Finland, France or Spain. The questions on money owed to the household, extra assets in managed accounts and the residual question on other financial assets are not asked in Finland. The question on investment attitudes is not asked in Finland or France.

**EMPLOYMENT**

The secondary labour status question (in addition to the main labour status) is not asked in Finland or the Netherlands. Occupation (ISCO) and sector of employment (NACE) are not collected in Italy. Type of contract is not collected in Finland. Time spent in the current main job is not collected in Finland and collected only for employees in Spain. Type of secondary employment is not collected in Finland. Total time spent in employment since the age of 16 is not collected in Finland or Italy. Expected retirement age is not collected in Finland and is only partially collected in Italy and Spain.

## PENSIONS AND LIFE INSURANCE POLICIES

The questions on public and occupational pension plans are not asked in Finland. The question on current % of gross earnings contributing to public pension plans is not asked in Austria, France, Italy or Spain and is asked only for one main plan in Belgium. The current value of accounts in public pension schemes is not collected in Austria, Belgium, Italy, Malta, the Netherlands or Spain. The current value of accounts in occupational pension schemes is not collected in Malta. The value of whole life insurance contracts is not collected in Finland. The question on the type of voluntary public pension plan (pension plan / whole life insurance contract) is not asked in Austria.

## INCOME

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

## INTERGENERATIONAL TRANSFERS, GIFTS

The whole questionnaire section on intergenerational transfers and gifts is left out in Finland and Italy.

The question on expected gifts and inheritances in the future is not asked in Spain or France. The question on from whom the gift/inheritance was received is not asked in Slovakia and is only partly asked in Spain.

## CONSUMPTION AND SAVING

The whole questionnaire section on consumption and saving is left out in Finland. In France, the questions on food expenditures are asked only to a one-third sub-sample of households. The amount spent on food outside the home is not collected separately, but rather provided together with amount spent for food at home in Italy and Spain. The question on the purpose of saving is not asked in France or Italy and is only partially asked in Spain. The question on the character of last 12 months' expenses (high/normal/lower) is not asked in Italy. The questions on the comparison of the balance between income and expenses and on the source of extra income to meet expenses in households with expenses above income are not asked in Italy. The question on ability to get financial assistance from friends or relatives is not asked in Italy or Spain.

## 10.3 COLLECTION OF THE NON-CORE ITEMS IN THE FIRST SURVEY WAVE

The following table provides an overview of non-core variables covered in one or more of the HFCS country files in Wave 1.

**Table 10.2 HFCS non-core variables collected in national surveys**

| | |
|---|---|
| **Demographics** | |
| RNA0100 Previous country of residence | France |
| RNA0200 Citizenship | France, Italy, Luxembourg |
| PNA0100 Field of study | Italy, Spain |
| PNA0200 Health | Italy, Spain |
| PNA0300 Siblings | Italy, France |
| PNA0600 Education of father/mother | Italy |
| PNA0700 Occupation of father | Spain, Portugal, France |
| PNA0701 Occupation of mother | Spain, Portugal, France |
| 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** | |
| HNB0810 HMR - year of construction | Italy, Spain, Portugal, Greece |
| HNB0910 External support for housing acquisition | Portugal, Germany |
| HNB0920 HMR/Imputed rent | Italy, Greece |
| HNB130$x HMR mortgages: institution you have loan with | Spain |
| HNB1700 Overpaying/voluntary step-up payments on HMR mortgages | Portugal |
| HNB1710 Monthly amount of extra voluntary payments on HMR mortgages | Portugal |
| HNB190$x Other properties: how property was acquired | Italy, Spain |
| HNB200$x Remaining other properties: renting out of property | Italy, Spain |
| HNB201$x Other properties: how much rent is collected | Italy, Spain |
| HNB2100 Renting out of additional properties (other than HMR + 3 OP) | Italy, Spain |
| HNB2110 How much rent is collected from additional properties (other than HMR + 3 OP) | Italy, Spain |
| HNB2300 Overpaying/voluntary step-up payments: loans on other properties | Portugal |
| HNB2310 Monthly amount of voluntary payments: loans on properties other than HMR | Portugal |
| HNB2700 Purchase of property or consumer durables | Spain, France |
| HNB2710 Types of purchased properties or consumer durables | Spain, France |

| | |
|---|---|
| HNB2720 Expenditure on buying properties and/or consumer durables | Spain |
| HNB2800 Sold properties or consumer durables | Spain |
| HNB2820 Amount received - sale of properties and/or consumer durables | Spain |
| HNB3000 Reasons for moving | Luxembourg, Portugal |
| **Other liabilities, credit constraints** | |
| HNC005$x Non-collateralised loans: nature of the lender | Spain |
| HNC0125 Late or missed payments on loans | Spain, Luxembourg, Portugal |
| HNC0126 Any outstanding overdue payments | Luxembourg, Portugal |
| HNC0200 Reasons for being refused credit | Spain, Portugal |
| HNC0210 Reasons for not applying for credit due to perceived credit constrain | Luxembourg, Portugal |
| **Private businesses, financial assets** | |
| HND010$x Businesses: year the business was started | Spain, France |
| HND020$x Businesses: last year's total business sales | Portugal, France |
| 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, Greece |
| HND1000 Market value by type of bond | Italy |
| HND2100 Managed accounts - description | France |
| HND3000x Largest asset in HH balance sheet | Luxembourg, Portugal, Belgium |
| HND3010 Portfolio shifts last two years? | Luxembourg, Portugal, Belgium |
| HND3020 Portfolio shifts last two years: money out | Luxembourg, Portugal, Belgium |
| HND3030 Portfolio shifts last two years: money in | Luxembourg, Portugal, Belgium |
| HND3040 Would not invest again? | Luxembourg, Belgium, Germany |
| HND3050 Assets HH would not invest again | Luxembourg, Belgium |
| HND3100 Net worth past two years | Luxembourg, Portugal, Belgium, Germany, France |
| HND3200 Net worth next two years | Luxembourg, Portugal, Belgium, Germany |
| **Employment** | |
| PNE0100 Seasonal employment | Portugal |
| PNE0110 Number of working weeks per year | Italy |
| PNE0200 Gross monthly income – main job (employees) | Spain |
| PNE0300 Gross monthly income from self-employment | Spain |
| PNE0500 Private-public organization | Italy, Luxembourg, Portugal, France |
| PNE0600 Number of employees – main employer | Spain, Luxembourg, Portugal |

| | |
|---|---|
| PNE0700 Hours worked – additional employment contracts (as an employee) | Italy, Spain |
| PNE0800 Gross monthly income from additional jobs | Spain |
| PNE1000 Looking for job | Spain, France |
| PNE2000 Former job title and description / ISCO | Spain, Germany, France |
| PNE2100 Time in former employment | Germany |
| PNE2200 Total time in full-time employment | Spain, Luxembourg |
| PNE2400 Number of different employers | Italy, Spain |
| PNE2700 Worsening of job conditions past two years | Portugal, Germany |
| PNE2800 Expected worsening of job conditions next two years | Portugal, Germany |
| **Pensions and life insurance** | |
| PNF040$x Public plans: years contributing | Italy |
| PNF050$x Public plans: expected age to receive benefits | Italy |
| PNF100$x Occupational plans: is employer contributing | Italy, Spain, France |
| PNF120$x Occupational plans: years contributing | Italy, Spain, France |
| PNF131$x Occupational plans: value of account | Italy, Spain, France |
| PNF180$x Occupational plans: expected age of collecting pension | Italy, France |
| PNF2000 Number of voluntary private pension plans | Italy, Spain, France |
| PNF210$x Type of voluntary pension plan | Italy, Spain, France |
| PNF220$x Voluntary pension plans: years contributing | Italy, Spain, France |
| PNF230$x Voluntary pension plans: contributions | Italy, Spain, France |
| PNF280$x Voluntary pension plans: age to start receiving payments | France |
| PNF290$x Voluntary pension plans: kind of payment at retirement age | Spain, France |
| PNF300$x Voluntary pension plans: value of account | Italy, Spain, France |
| PNF310$x Whole life insurance policy: cash value | Spain, France |
| PNF311$x Voluntary plans - expected age to collect pension | Italy, Spain, France |
| PNF3600 Has private health insurance | Italy, Spain |
| PNF3610 Monthly payments for health insurance policy(ies) | Italy |
| **Income** | |
| PNG0110 Net employee income | Italy, Greece |
| PNG0210 Net self-employment income | Italy |
| PNG0310 Net income from public pensions | Italy |
| PNG0410 Net income from private and occupation pension plans | Italy |
| PNG0510 Net income from unemployment benefits | Italy |
| HNG0110 Net income from regular social transfers | Italy |

| | |
|---|---|
| HNG0210 Net income from regular private transfers | Italy |
| HNG0310 Net rental income from real estate property | Italy |
| HNG0410 Net income from financial investments | Italy |
| HNG0510 Net income from private business other than self-employment | Italy |
| HNG0610 Net income from other sources | Italy |
| HNG0710 Income taxes and social contributions | Italy, Finland |
| **Intergenerational transfers, gifts** | |
| HNH0500 Substantial gift made to children/other people outside household | Portugal, France |
| HNH0600 Who was the beneficiary of the gift | Portugal, France |
| HNH0700 Year donation was made | Portugal, Belgium, France |
| HNH0800 How much was donation made worth | Portugal, Belgium, France |
| **Consumption and saving** | |
| HNI0100 Expenditure on utilities | Portugal, France |
| HNI0210 Expenditure on regular payments | Portugal |
| HNI0300 Total consumption expenditure | Spain, Portugal, France |
| HNI0700 More or less savings in the next year | Luxembourg, Germany |
| HNI0800 General price expectations | Luxembourg, Germany |
| **Payment habits (non-core section)** | |
| HNJ1100 Any debit or/and ATM cards (y/n) | Italy, Spain |
| HNJ1300 Frequency of cash withdrawals in ATMs | Spain |
| HNJ1400 Use of direct debit (y/n) | Spain |
| HNJ1500 Type of payments by direct debit | Spain |
| HNJ1600 Reasons for not using direct debit | Spain |
| HNJ1700 Frequency of bank transfers | Spain |
| HNJ1800 Payments by bank cheques | Spain |
| HNJ2000 Any payments received by credit transfer | Italy, Spain |
| HNJ2200 Own credit or store cards | Italy |
| HNJ2300 No. of credit/store cards | Italy |
| HNJ2500 Total monthly payment on all these cards | Spain |
| HNJ2800 Ever used other means of payment | Spain |
| HNJ2900 Link used for info or payments | Spain |
| HNJ3000 How frequently uses other means of payment | Spain |
| HNJ3200 Any household member use the internet | Spain |
| HNJ3800 Cash at home to meet normal needs | Spain |

## 10.4 COMPARING HFCS AND NATIONAL ACCOUNTS AND ASSESSING CONSISTENCY BETWEEN CONCEPTS AND DEFINITIONS

This chapter analyses the consistency and comparability between the balance sheet data from the Eurosystem Household Finance and Consumption Survey (HFCS) and the National Accounts data. Due to significant differences related to methodology, coverage, etc. between the

two sources, the comparison of wealth items requires special attention. After describing some general differences between the two sources, the consistency assessment is done at the level of individual survey variables and National Accounts items. This chapter summarises the main challenges; for a more comprehensive analysis of the comparability between micro and macro sources, as well as initial conclusions based on country data, see Kavonius and Törmälehto (2010) and Honkkila and Kavonius (2012). The references mentioned above also assess the comparability between income concepts. Chapter 10.5 evaluates the comparability between the income data of the HFCS with the EU-Statistics on Income and Living Conditions (EU-SILC), which is a source with more comparable definitions and data collection methods.

Information on household income and wealth can be found in two sources: microdata coming from surveys, and macro series collected for the compilation of National Accounts. Coming from different traditions and addressing different purposes, the micro and macro approaches have developed quite independently. Survey microdata aim at analysing income and wealth distributions, as well as at comparing income, wealth and debt across different sub-populations. Household-level data allow important insights into the economic behaviour of households that cannot be achieved 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. 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 tries to minimise bias in the estimates for the economy as a whole, as well as to minimise statistical discrepancies within the system. Thus, 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.

A necessary condition for micro-macro comparisons to be meaningful is that concepts and definitions underlying the two sources are consistent. This issue is linked to at least three conditions:

i) The boundaries of the household sector

ii) The existence and definitions of items to be included in the various measures of wealth

iii) Valuation of assets and reference periods

### 10.4.1 THE BOUNDARIES OF THE HOUSEHOLD SECTOR

According to the European System of Accounts (ESA) definition, the household sector (S.14) includes consumer households and producer households. Sector S.15 includes Non-profit institutions serving households (NPISHs), which are private, non-market producers that provide goods or services to households for free or at prices that are not economically significant, such as churches and religious societies, sports and other clubs, trade unions and political parties. While National Accounts should ideally provide separate figures for the Household sector and NPISHs, most usually this is not the case. Furthermore, NPISHs have been observed to possess

significant amounts of wealth. Therefore, the micro-macro comparison would require deducting their amounts from National Accounts. Additionally, most household surveys (and the HFCS in particular) do not cover persons living in institutions.

Moreover, the distinction between producer households (to be classified within the household sector) and quasi-corporations (to be classified within the non-financial corporations sector) in National Accounts is relevant, because it affects gross wealth and the composition of the household balance sheet. If such businesses are considered part of the household sector, their (financial and non-financial) assets and liabilities are also part of the household balance sheet. Conversely, if they are classified as separate institutional units, the balance sheet of the owner household only records a (net) participation in equity, i.e. a financial asset. In turn, when comparing survey data with National Accounts, the distribution of real and financial assets as well as gross assets and liabilities of households register will be different, while the level of total net wealth remains unaffected. Unfortunately, ESA does not set any clear-cut condition for deciding whether a certain unincorporated business owned by households should be classified as a separate institutional unit, i.e. as a producer household or as a quasi-corporation, and national practices vary a lot. The main determinants are the number of employees and the legal form of the business. In the HFCS, self-employment businesses' main activity, number of employees and legal form are collected. Theoretically, this information could make it possible to identify producer households in some (not all) countries and thereby define a household sector more comparable with National Accounts.

### 10.4.2 COMPARABILITY OF WEALTH ITEMS

Tables 10.3 and 10.4 show the correspondence of balance sheet items between the HFCS and National Accounts. Table 10.3 has the information on real wealth. Of non-financial assets, the value of dwellings (AN.1111) in the National Accounts is included in two HFCS items, "The value of the Household main residence" and "The value of other properties", as long as the type of other property belongs to the category of dwellings. Additionally, the value of land (AN.211) is included in the values of the household main residence and other properties in the HFCS. Part of the other properties in HFCS can be categorised under the National Accounts item other buildings and structures (AN.1112), as long as that property belongs to this class of non-financial assets. Information on the type of other properties is included only for the three most important properties per household. For remaining properties, it is not possible to determine whether they should be classified as dwellings or other buildings and structures.

The value of valuables is collected in the HFCS, and by and large corresponds to the National Accounts item "Valuables" (AN.13). The value of cars and other vehicles is collected in the HFCS, but is not included as a separate item in National Accounts. They are included only in the memorandum item "Consumer durables", unless some of them have been identified as being used in the production activities.

**Table 10.3 Correspondence table- household real wealth in HFCS and National Accounts**

| ESA95 Code | National Accounts | HFCS | Remarks |
|---|---|---|---|
| **AN.1 Produced assets** | | | |
| **AN.11 Fixed assets** | | | |
| **AN.111 Tangible fixed assets** | | | |
| AN.1111 | Dwellings | Household main residence, other properties | Other properties reported as dwellings. |
| AN.1112 | Other buildings and structures | Other properties | Other properties reported as other buildings and structures. |
| AN.1113 | Machinery and equipment | Value of self-employment business | Included in business value that is measured in net terms. Part of this item is recorded as financial wealth of the household sector in NA. |
| AN.1114 | Cultivated assets | Other properties | Farm and land are categories for other property types. This item cannot, however, be measured separately. |
| **AN.112 Intangible fixed assets** | | | |
| AN.12 | Inventories | Value of self-employment business | Included in business value that is measured in net terms. Part of this item is recorded as financial wealth of the household sector in National Accounts. |
| AN.13 | Valuables | Valuables | |
| **AN.2 Non-produced assets** | | | |
| AN.211 | Land | Included in values of the household main residence and other properties | |
| AN.212-214 | Subsoil assets, etc. | | |
| AN.22 | Intangible non-produced assets | | Included in business value that is measured in net terms. Part of this item is recorded as financial wealth of the household sector in National Accounts. |
| | Consumer durables (memorandum item) | Cars and other vehicles | Included only if not used in production of self-employment business. |

Probably the most significant issue in the comparison of real wealth totals between the HFCS and National Accounts is the treatment of business wealth and the sector delineation of especially self-employment businesses. In the HFCS, all assets and liabilities of self-employment businesses (where at least one household member works for the business) are recorded as real wealth, while investments in the equity of non-self-employment businesses (whether publicly traded or not) are classified as financial assets. National Accounts record all households' property rights on entities involved in production (of course, only for those which are considered as institutional units separate from the households – see section 1.1) as equity participations (i.e. financial assets), whereas household participations in unincorporated enterprises classified as producer households are spread over the various asset and liability items in the household balance sheet (buildings, land, machinery, inventories, loans, etc.). The latter assets and liabilities cannot be distinguished from those corresponding to the household itself.

Indeed, self-employment business wealth is reported in the HFCS in net terms only. While the treatment of business wealth in National Accounts and in the HFCS has no impact on the comparability of total net wealth, it affects the totals of both gross assets and liabilities. Moreover, National Accounts items related to entrepreneurial activities of the household sector, such as machinery and equipment, inventories and goodwill, cannot be separated in the HFCS data. However, the value of these assets should be included in the net value of self-employment business wealth.

For several financial assets collected in the HFCS, there are corresponding items in the National Accounts (see table 10.4). National Accounts items F22+F29 "Deposits" correspond to the HFCS items "Sight accounts" and "Savings accounts", National Accounts item F33 "Securities other than shares except financial derivatives" to the HFCS item "Bonds", National Accounts item F511 "Quoted shares" to the HFCS item "Publicly traded shares", and the National Accounts item F52 "Mutual fund shares" to the HFCS item "Investments in mutual funds". For these assets, the definitions are identical or very similar in both data sources.

Information on currency (item F.21 in National Accounts) held by households is not collected in the HFCS. The HFCS item "Managed accounts" is not considered a separate financial instrument in National Accounts. Various assets held through this kind of account are assigned to the corresponding category in the household's balance sheet. Financial derivatives (F.34) are included as a separate item in the National Accounts, but not in the HFCS. Financial derivatives such as options, futures or index certificates should be covered by the survey variable "other financial assets". Loans granted by households are an item both in the National Accounts (F.4) and in the HFCS. However, in National Accounts, loans between households are not included in this item.

**Table 10.4 Correspondence table- household financial wealth in HFCS and National Accounts**

| ESA95 Code | National Accounts | HFCS | Remarks |
|---|---|---|---|
| **F.2 Currency and deposits** | | | |
| F.21 | Currency | | Not collected in HFCS. |
| F.22 | Transferable deposits | Sight accounts | |
| F.29 | Other deposits | Savings accounts | |
| **F.3 Securities other than shares** | | | |
| F.33 | Securities other than shares, excluding financial derivatives | Bonds | |
| F.34 | Financial derivatives | Other financial wealth | Not a separate item in HFCS. |
| F.4 | Loans | Amount owed to household | Not fully comparable, loans between households missing from National Accounts. |
| **F.5 Shares and other equity** | | | |
| F.511 | Quoted shares | Publicly traded shares | |
| F.512 | Unquoted shares | Investment in non-self-employment not publicly traded shares | In the HFCS, classification is based on household activity in the enterprise. National Accounts value includes assets that are classified as real wealth in the HFCS. Not fully comparable. |
| F.513 | Other equity | | |
| F.52 | Mutual funds shares | Investments in mutual funds | |
| F.6 | Insurance technical reserves | Voluntary pension/whole life insurance schemes | Only whole life insurance collected in HFCS. |
| F.7 | Other accounts | Other financial assets | Not comparable, different definitions. |
| | | Managed accounts | Included in values of other financial assets in National Accounts. |
| F.L | Liabilities | Several variables depending on the collateral used to take the loan | |

The National Accounts concept of insurance technical reserves (F.6) may be interpreted as the functional equivalent of pension wealth in the HFCS. However, the HFCS net wealth concept includes only the current termination value of (funded) private pension plans, i.e. excluding

public and occupational pension plans and social security funds, while part of these assets (namely participations in plans other than social security schemes) are included in the National Accounts.

Loans in the HFCS are broken down by collateral (i.e. the household's main residence [HMR], other real estate properties and other / non-collateralised). Breakdowns by purpose of the loan, type of interest, maturity, etc. are also included in the survey for the most important household loans within each category (collateralised by the HMR, collateralised by other properties or uncollateralised). Loans in National Accounts are broken down according to their maturity. Loans are created when creditors lend funds to debtors, either directly or through brokers, which are either evidenced by non-negotiable documents or not evidenced by documents. The definitions of liabilities are similar in the National Accounts and the HFCS, with the exception of liabilities of self-employment businesses that have been discussed above.

### 10.4.3 VALUATION OF ASSETS AND REFERENCE PERIODS

The valuation of assets (particularly of real assets) in the HFCS is based on the household self-assessment, while National Accounts data are at estimated market values. In some instances, National Accounts data are a residual of the estimates of other institutional sectors, and the results do not always perform well when compared to aggregates from administrative records[42].

The valuation method chosen for real estate in the HFCS was a conscious decision based on several reasons: first, standardised publicly available prices for real estate property are not generally available, and less so at times at which the number of transactions is relatively low. Furthermore, it is the household's self-perception about the value of its properties that determines the economic and financial decisions that they take. While such different valuation methods should theoretically lead to very similar outcomes, this difference should be taken into account when comparing the results, especially during periods of significant volatility in the markets or during times when markets are thin and market prices thus difficult to gauge.

The reference periods of different HFCS country data vary from 2008 to 2010, although for most euro area countries the reference year for balance sheet items is 2010. Therefore, it is crucial to make the comparison to corresponding periods of National Accounts data. For income items, the survey reference period is in most cases an entire calendar year, making it perfectly comparable with the National Accounts. However, in certain limited cases the reference period for wealth is the last day of the year. To minimise recall bias, most wealth surveys measure the values for households' balance sheet items at the time of the interview. This can have an effect on comparability when asset values are changing rapidly.

[42] For instance, in Spain, the Financial Accounts estimate of household debt in the last quarter of 2008 was 15% higher than the correspondent figure from banks' balance sheets.

### 10.4.4 OVERALL DESCRIPTION OF THE RESULTS FOUND IN THE INTERNAL CALCULATIONS CARRIED OUT BY THE ECB

There are only limited comparable data sources available at the macro level on real assets in the euro area countries, as well as for non-profit institutions serving households. For financial assets and liabilities in sector S.14+S.15 (households including NPISHs), as well as for income, data availability is much better. With the data sources available, the ECB has performed some internal tests to assess comparability between the HFCS and National Accounts data. In the comparisons performed, even for a limited number of countries (nine), several estimation methods had to be applied in the production of National Accounts figures for real assets.

To give an example of the kind of coherence and plausibility checks performed against external benchmarks, the table below presents a comparison between per capita HFCS assets, liabilities and net wealth with the same magnitudes in National Accounts. This comparison covers real assets, as well as gross and net wealth of ten countries for which macro level data on real assets were available, either through Eurostat or directly from national sources. For four additional countries, levels of financial assets and liabilities are compared. In the comparison between National Accounts data, it must be highlighted that neither the wealth definitions nor populations covered are strictly comparable.

**Table 10.5 Comparison between per capita HFCS and per capita National Accounts**

| | HFCS (per capita) | | | | |
|---|---|---|---|---|---|
| | Real assets | Financial assets | Total assets | Liabilities | Net wealth |
| Belgium | 113,000 | 46,000 | 160,000 | 13,000 | 147,000 |
| Germany | 86,000 | 23,000 | 109,000 | 13,000 | 95,000 |
| Greece | 56,000 | 4,000 | 60,000 | 5,000 | 56,000 |
| Spain | 108,000 | 12,000 | 121,000 | 12,000 | 109,000 |
| France | 93,000 | 22,000 | 115,000 | 11,000 | 104,000 |
| Italy | 102,000 | 11,000 | 113,000 | 5,000 | 109,000 |
| Cyprus | 247,000 | 22,000 | 267,000 | 26,000 | 243,000 |
| Luxembourg | 283,000 | 36,000 | 319,000 | 33,000 | 286,000 |
| Malta | 115,000 | 18,000 | 133,000 | 4,000 | 128,000 |
| Netherlands | 84,000 | 30,000 | 114,000 | 37,000 | 77,000 |
| Austria | 111,000 | 22,000 | 133,000 | 8,000 | 125,000 |
| Portugal | 55,000 | 8,000 | 63,000 | 6,000 | 57,000 |
| Slovenia | 57,000 | 3,000 | 60,000 | 2,000 | 58,000 |
| Slovakia | 27,000 | 2,000 | 30,000 | 1,000 | 28,000 |
| Finland | 81,000 | 14,000 | 95,000 | 17,000 | 78,000 |
| | **National accounts (per capita)(*)** | | | | |
| | Real assets | Financial assets | Total assets | Liabilities | Net wealth |
| Belgium | 94,000 | 80,000 | 174,000 | 17,000 | 157,000 |
| Germany(***) | 101,000 | 54,000 | 155,000 | 19,000 | 137,000 |

| | | | | | |
|---|---|---|---|---|---|
| Greece | | 24,000 | | 11,000 | |
| Spain (****) | 129,000 | 34,000 | 162,000 | 20,000 | 142,000 |
| France | 112,000 | 58,000 | 170,000 | 17,000 | 153,000 |
| Italy | 97,000 | 56,000 | 153,000 | 12,000 | 141,000 |
| Cyprus | | 49,000 | | 27,000 | |
| Luxembourg | | 92,000 | | 37,000 | |
| Malta | | 36,000 | | 9,000 | |
| Netherlands | 96,000 | 94,000 | 190,000 | 44,000 | 146,000 |
| Austria(***) | 103,000 | 58,000 | 161,000 | 19,000 | 141,000 |
| Portugal | | 35,000 | | 15,000 | |
| Slovenia(***) | 52,000 | 16,000 | 68,000 | 5,000 | 63,000 |
| Slovakia(***) | 32,000 | 7,000 | 39,000 | 3,000 | 36,000 |
| Finland(***) | 80,000 | 39,000 | 119,000 | 20,000 | 99,000 |

(*) Source: Eurostat, National Bank of Belgium, Banco de España, Banca d'Italia, estimates by the Banque de Luxembourg on the basis of banking statistics and internal ECB calculations

(**) Excluding currency

(***) Value of land estimated as a proportion of dwellings based on data available for FR and NL.

(****) Real assets include only housing stock and land underlying buildings.

**HFCS / National accounts, %**

| | Real assets | Financial assets | Total assets | Liabilities | Net wealth |
|---|---|---|---|---|---|
| Belgium | 121% | 58% | 92% | 76% | 94% |
| Germany | 85% | 42% | 70% | 71% | 70% |
| Greece | | 18% | | 41% | |
| Spain | 84% | 37% | 75% | 61% | 76% |
| France | 83% | 39% | 68% | 67% | 68% |
| Italy | 105% | 20% | 74% | 40% | 77% |
| Cyprus | | 45% | | 96% | |
| Luxembourg | | 39% | | 90% | |
| Malta | | 53% | | 46% | |
| Netherlands | 87% | 32% | 60% | 84% | 53% |
| Austria | 108% | 38% | 82% | 41% | 88% |
| Portugal | | 22% | | 42% | |
| Slovenia | 109% | 21% | 88% | 38% | 92% |
| Slovakia | 83% | 35% | 75% | 38% | 78% |
| Finland | 101% | 37% | 80% | 88% | 78% |

As expected, given the lesser coverage of wealth items and a more limited definition of the household sector, the net wealth levels are lower in the HFCS than in the National Accounts. Mostly because of the delineation of self-employment businesses, the average values of real wealth are relatively high in the HFCS, in some cases even higher than in the National Accounts. The average values of financial assets, on the other hand, are relatively low. In overall wealth for Italy and Slovenia, the low values for financial assets in the HFCS are offset by the relatively high values of real assets. Currency was excluded from the National Accounts figures on financial wealth, since it is not part of the HFCS financial wealth. The comparison between liabilities only includes the National Accounts item "Loans", comparable with the

HFCS definition, and only the net value of the item "Other accounts receivable/payable" is included in the net wealth figures of National Accounts.

Part of the lower values of financial assets and liabilities in the HFCS, as well as cross-country differences, are explained by the conceptual issues described earlier. The National Accounts figures include the whole sector S.14, including NPISHs. In countries that report households separately from NPISHs, it has been observed that the financial assets of NPISHs account on average for 5% of total financial assets of the sector S.14+S.15. There are, however, significant differences between countries and individual assets. The share of liabilities for NPISHs is much smaller, 1-2% in all cases observed.

All self-employment business wealth is classified as real assets in the HFCS. A different delineation of self-employment business assets would have a considerable impact on the financial wealth levels and on the comparability to National accounts especially in Cyprus[43], Malta and Portugal. However, the impact of such re-classification varies a lot between countries[44]. Additionally, the differences in the definition of pension wealth might have a particularly substantial impact on comparability.

The significance of the differences in definitions and concepts that cannot be corrected by simply excluding or re-classifying items in either data source is difficult to evaluate. The different purposes of the two sources might be another cause for divergences. Even though wealthy households are oversampled in the HFCS, household surveys are generally not able to cover the ultimate top of the wealth distribution (not least because of confidentiality reasons). Wealth, especially financial wealth, is very unequally distributed and the share of the highest percentiles of the distribution in total wealth can be substantial. Consequently, if the wealthiest of the wealthy are not captured by the survey, it will influence the wealth totals and means. Nevertheless, the impact on the most important indicators produced by the survey for other parts of the wealth distribution should be limited.

## 10.5 COMPARISON OF INCOME DATA BETWEEN THE HFCS AND EU-SILC

EU-SILC provides a useful benchmark for comparing income data of the HFCS. Unlike in the case of National Accounts, EU-SILC, being a household survey, is conducted for similar purposes and uses data collection methods similar to those of the HFCS. It should be acknowledged, though, that the HFCS aims at maximising the efficiency of the estimates of the wealthiest households, while the main target of the EU-SILC is low income households. This

[43] Cyprus, as a small economy has a large number of self-employed/small businesses. However, there is not a clear line for the classification into S11 (Non-financial corporations) and S14 (Households), i.e. a small business owned by a household may be classified as a property of the household and not as a non-financial corporation. In principle, the balance sheet of financial assets in NA of Cyprus includes the values of self-employment businesses other than sole proprietorships. However, the 2010 NA figures for unquoted shares and other equity (F512-513) are estimates based on a 2001 survey of businesses and are hardly comparable with HFCS values.

[44] Self-employment business wealth of legal forms other than sole proprietorships is usually classified as financial assets of the household sector in National Accounts. The value of such business wealth in the HFCS is 230% of other financial assets in Cyprus, 60% in Malta and 80% in Portugal, but only 2% and 8% in the Netherlands and Greece respectively.

leads to different sampling strategies in these surveys. Both surveys share, to a large extent, identical concepts and definitions of the target population and of income. That said, both some general and some country-specific differences in concepts and methodologies should be noted. Given the differences and common challenges in data production methodologies, one should not consider either of the two surveys the absolute benchmark for income data. Nevertheless, similar results from two household surveys sharing a wide range of similar methodologies should provide positive signals for the quality of both surveys.

The definitions of household and the target population are identical in both surveys. However, in Italy the EU-SILC definition of private households ("Cohabitants related through marriage, kinship, affinity, patronage and affection"). is different from the one used in other countries and in the HFCS. In Austria, the target population of EU-SILC includes only households living in a dwelling officially registered in the Austrian population register as a main residence, while the HFCS target population also includes households living in dwellings which are not registered as a main residence.

Some differences in the data collection methods can be observed between EU-SILC and HFCS. In seven countries, the main data collection method was the Computer Assisted Personal Interview (CAPI) for both EU-SILC and the HFCS. In Finland, both surveys use Computer Assisted Telephone Interviews (CATI). Of countries collecting data via CAPI in the HFCS, in Greece, Italy, Luxembourg, Slovenia and Slovakia the dominant data collection method for EU-SILC was Paper-and pencil Assisted Personal Interviews (PAPI); in Germany it was self-administered interviews. In Cyprus, the main data collection method for EU-SILC was CAPI, and for HCFS, PAPI. In the Netherlands, CATI was applied in EU-SILC, while HFCS data are collected with web-based interviews. However, in Finland and France most income data are derived from administrative sources for both surveys, while in the Netherlands and Slovenia administrative sources are used for EU-SILC only.

In the HFCS, the income concept is gross income, i.e. taxes, social contributions and other transfers paid by households are not deducted from the income totals. Consequently, comparisons with external sources should only be made to similar income concepts, and not to after-tax income (disposable income). Data from EU-SILC enables a comparison to a concept of gross income that is identical with the HFCS one, with the exception of income from private use of a company car that is not included in the HFCS. Table 10.6 shows the correspondence between individual income items collected in the two surveys. For most individual items, EU-SILC definitions were applied as such to the HFCS, although some differences that are explained in the table below remain. Data on social transfers in EU-SILC are collected in a more detailed manner, while financial income is more detailed in the HFCS.

**Table 10.6 Correspondence table- household gross income in HFCS and EU-SILC**

| EU-SILC | HFCS | Remarks |
|---|---|---|
| Employee cash or near cash income | Employee income | |
| Income from private use of company car | | Not included in HFCS |
| Cash benefits or losses from self-employment | Self-employment income | |
| Old-age benefits<br>Survivors' benefits<br>Disability benefits | Income from public pensions | |
| Pension from individual private plans | Income from private and occupational pensions | Pensions from mandatory employer-based schemes included in public pensions in EU-SILC |
| Unemployment benefits | Income from unemployment benefits | Severance and termination payments and redundancy compensation included in other income in the HFCS. |
| Sickness benefits<br>Education related allowances<br>Family/Children related allowances<br>Social exclusion not elsewhere classified<br>Housing allowances | Income from regular social transfers | |
| Regular inter-household cash transfer received | Income from regular private transfers | |
| Income from rental of a property or land | Rental income from real estate property | |
| Interest, dividends, profits from capital investment in an unincorporated business | Income from financial investments<br>Income from private business other than self-employment | |
| Income received by people under age 16 | Income from other income source | Personal level variables, such as employee or self-employment income asked in HFCS only for persons 16 and over. |

Table 10.7 below provides a comparison of the median household gross income between HFCS and EU-SILC. The coherence between the figures is very good, especially taking into account some differences in definitions.

Table 10.7 Comparison of median income in EU-SILC and in the HFCS

| Country | Median gross income HFCS, € | Median gross income EU-SILC, € | HFCS, |
|---|---|---|---|
| Belgium | 34,000 | 35,000 | 97% |
| Germany | 33,000 | 33,000 | 98% |
| Greece | 22,000 | 24,000 | 92% |
| Spain | 25,000 | 26,000 | 96% |
| France | 29,000 | 36,000 | 81% |
| Italy | 26,000 | 31,000 | 85% |
| Cyprus | 32,000 | 34,000 | 94% |
| Luxembourg | 65,000 | 66,000 | 98% |
| Malta | 22,000 | 22,000 | 97% |
| Netherlands | 41,000 | 43,000 | 95% |
| Austria | 32,000 | 41,000 | 78% |
| Portugal | 15,000 | 17,000 | 86% |
| Slovenia | 18,000 | 23,000 | 78% |
| Slovakia | 11,000 | 12,000 | 93% |
| Finland | 36,000 | 36,000 | 101% |

## 10.6 STATISTICAL DISCLOSURE: ADDITIONAL INFORMATION

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

## 10.6.1 VARIABLE DELETION AND RECODING

**Table 10.8 Variable deletion and recoding**

| | |
|---|---|
| **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, which are copied without modification to H file.<br>Starting in wave 2, the following variables will also be copied without modification: SA0110 past household ID (panel), SA0210 vintage of past interview |
| **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 MT. |
| RA0400 country of birth | Recoded to "local"/OEA/OEU/OTH where local is the ISO two digit code for the survey country[45]. 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 FI, IT, MT, 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 and PT. |
| HB0200 length in residence | Top-coded at 85. Binding in BE, ES, IT, MT and PT. |
| HB0700 year of acquisition | Bottom-coded at 1925.[46] Binding in ES, IT, LU, MT and PT. |
| **Financial assets** | |
| HD050$x_B number of employees | Created from HD050$x. Brackets: 0, [1;2], [3;9], [10,+∞). Applied in BE, DE, LU, MT, and PT. |
| HD1910 specification of assets | Deleted (verbatim answer)[47] |
| **Employment** | |
| PE1000 time in main job | Top-coded at 73.[46] Binding in ES.<br>Top-coded at 69 applied in PT. |

[45] 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.

[46] 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., 2010-85). 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 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 1938.

[47] Verbatim answers can 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.

| | |
|---|---|
| **Pensions** | |
| PF0300 years contributing | Top-code at 73. [46] Binding in BE and ES. |
| **Income** | |
| HG0620 sources of other income | Deleted (verbatim answer)[47] |
| **Inheritances** | |
| HH020$x year received | Bottom-code at 1925. [46] Binding in BE, MT, PT. |
| **Interview closure** | |
| HP0100 items difficult | Deleted |
| HP0200 items missed | Deleted |
| HP0300 interviewee additions | Deleted |
| **Paradata** | |
| HR0100 to HR1100 | Deleted |
| HR1300 to HR1600 | Deleted |

### 10.6.2 TOP-CODING AND GROUPING

**Table 10.9 Top-coding and grouping**

| | |
|---|---|
| HB1010 | number of HMR mortgages. Top-coded at 3 in BE. |
| HB2410 | number of properties. Top-coded at 6 in LU, MT, and PT. Top-coded at 12 in BE. |
| HB250x | type of other properties. Categories Office, Hotel, Farm and Industrial buildings/Warehouse merged with Other in MT. |
| HB3010 | number of other mortgages. Top-coded at 4 in BE. |
| HB4310 | number of cars. Top-coded at 3 in BE. Top-coded at 4 in LU, MT, and PT. |
| HB4510x | number of other vehicles (motorbikes, trucks, vans, planes, boats/yachts, other). 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. |
| 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. |
| PF0610 | number of retirement plans. Top-coded at 5 in PT |
| HH060$x | source of inheritance/gift |
| PNA0300 | number of siblings |

# REFERENCES

**Methodological Publications about the country surveys participating in the HFCS**


Albacete, N., Lindner, P., Wagner, K., and Zottel, S. (2012), "Eurosystem Finance and Consumption Survey 2010. Methodological Notes in Austria", *Monetary Policy & the Economy* Q3/12 – Addendum, OeNB.

Banca d'Italia (2012), "Household Income and Wealth in 2010", *Supplements to the Statistical Bulletin, Sample Surveys*, Number 6.

Bundesbank (2013), forthcoming

Bover, O. (2011), "The Spanish survey of household finances (EFF): description and methods of the 2008 wave", *Documentos Ocasionales*, Nº 1103, Banco de España.

Costa, S. and L. Farinha (2012) "Household Finance and Consumption Survey: methodology and main results", *Occasional Paper* No.1, Banco de Portugal, forthcoming.

Faiella, I. (2008), "Accounting for sampling design in the SHIW", *Temi di discussione (Economic working papers)* 662, Bank of Italy, Economic Research and International Relations Area.

Faiella, I., and R. Gambacorta (2007), "The Weighting Process in the SHIW", *Temi di discussione (Economic working papers)* 636, Bank of Italy, Economic Research and International Relations Area.

Mathä, T., A. Porpiglia and M. Ziegelmeyer (2012), "The Luxembourg Household Finance and Consumption Survey (LU-HFCS): Introduction and Results", *Working paper Banque Centrale du Luxembourg,* N°73.

National Bank of Belgium (2013), forthcoming

Teppa, F., and C. Vis (2012), "The CentERpanel and the DNB Household Survey: Methodological Aspects", *DNB Occasional Studies* Vol.10/No.4.

Tzamourani, P. (2012), "Coherence assessment of the Greek HFCS with comparable data sources". *Q2012 - European Conference on Quality in Official Statistics, Athens, May2012*. (http://www.q2012.gr/articlefiles/sessions/8.3tzamourani_q2012_revised.pdf)


**Other references**


Altissimo, F., E. Georgiou, T. Sastre, M. Valderrama, G. Sterne, M. Stocker, M. Weth, K. Whelan and A. Wilman (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.*

Barceló, Cristina (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 86(4), 948-955.

Berben, R.P., K. Bernoth and M. Mastrogiacomo (2006), "*Households' Response to Wealth Changes: Do Gains or Losses Make a Difference?*" DNB Working Paper No. 90.

Biancotti, C.; Kennickell, A.; Sánchez Muñoz, C. (2009), "*Decentralised multi-country imputation: the case of the euro area Household Finance and Consumption Survey*" Supporting Paper 25, UNECE Conference of European Statisticians (http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2009/wp.25.e.pdf)

Bledsoe, R, and .Fries (2002), "Editing the 2001 Survey of Consumer Finances", Annual Meeting of the American Statistical Association, Joint Statistical Meetings, New York, New York, August 11–15, 2002.

Browning, Crossley and Weber (2003) *Asking Consumption Questions in General Purpose Surveys.* The Economic Journal, Vol. 113, No. 491, Features (Nov., 2003), pp. F540-F567.

De Leeuw, Edith 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, Don A. and Leah Melani Christian (2005): Survey Mode as a Source of Instability in Response Across Surveys. *Field Methods*, 17,1, pp. 30-52.

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

Engelhardt, G. V. (1996), "*House Prices and Home Owner Saving Behavior*", Regional Science and Urban Economics, 26(3/4), pp. 313-336.

Eurostat (2011): *EU-SILC, Description of Target Variables: Cross-sectional and Longitudinal*. European Commission, Eurostat.

Expert Group on Household Income Statistics, *The Canberra group Final Report and Recommendations* (2001), http://www.lisproject.org/links/canberra/finalreport.pdf

Faiella, Ivan (2010): "The Use of Survey Weights in Regression Analysis. Banca d'Italia Working Paper No. 739.

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

Groves, R.M. and E. Peytcheva (2008): The Impact of Non-response Rates on Non-response Bias. *Public Opinion Quarterly, Vol. 72, Issue 2, pp. 167-189.*

Horvitz, D.G. and D.J. Thompson (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 1. ECB, http://www.ecb.int/home/pdf/research/hfcn/core_questionnaire.pdf

Household Finance and Consumption Network: *HFCS core and HFCS non-core variables*, ECB, http://www.ecb.int/home/pdf/research/hfcn/core_output_variables.pdf

Household Finance and Consumption Network (2008a): Household Finance and Consumption Survey: Modalities for Implementation. ECB. (http://www.ecb.int/home/pdf/research/hfcn/hfcnImplementation.pdf)

Household Finance and Consumption Network (2008b): "*Imputation and Data Editing*" HFCN Implementation Documents. (http://www.ecb.europa.eu/home/pdf/research/hfcn/editing_imputation.pdf)

Household Finance and Consumption Network (2009): "*Oversampling the wealthy: eye for an eye, euro for a euro"*. HFCN Implementation Documents. (http://www.ecb.europa.eu/home/pdf/research/hfcn/Oversampling.pdf)

Household Finance and Consumption Network (2009): "*Survey Data on Household Finance and Consumption Research Summary and Policy Use"*. European Central Bank Occasional Paper Series Number 100 (http://www.ecb.europa.eu/pub/pdf/scpops/ecbocp100.pdf)

Household Finance and Consumption Network (2011): *Construction of Survey Weights for the HFCS*. ECB.

Jäckle, Annette; Caroline Roberts and Peter Lynn (2006): "Telephone Versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes". *ISER Working Paper* 2006-41.

Kennickell, Arthur (1991): *Imputation of the 1989 Survey of Consumer Finances: Stochastic Relaxation and Multiple Imputation*.

Kennickell, Arthur (1998): *Multiple Imputation in the Survey of Consumer Finances*.

Kennickell, Arthur 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, Arthur 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, Arthur B. (2007), "The Role of Over-sampling of the Wealthy in the Survey of Consumer Finances", Federal Reserve Board, Survey of Consumer Finances Working Papers, July 2007.

Kennickell, A. and J. Lane (2007) *Measuring the Impact of Data Protection Techniques on Data Utility Evidence from the Survey of Consumer Finances*, FRB Working Paper.

Lehtonen, R. and E. Pahkinen (2004), *Practical Methods for Design and Analysis of Complex Surveys 2nd edition*, John Wiley and Sons, Statistics in Practice.

Little, R. and Rubin, D. (2002), *Statistical Analysis with Missing Data*, 2nd ed. (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., A. Saïdi and R. Pettapiece (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.

Osiewicz, M. and S. Pérez-Duarte (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, http://goo.gl/YP9vC

Pérez-Duarte S., C. Sánchez Muñoz, and V.-M.Törmälehto*: 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, May 4-6 2010. http://www.ecb.int/home/pdf/research/hfcn/WealthSurveys.pdf

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

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

Rao, J. N. K, C. F. J. Wu and K. Yue (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.

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

Rubin, Donald B. (1987): *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons inc.

Rubin, Donald B. (1996), "Multiple Imputation After 18+ Years", *Journal of the American Statistical Association*, Vol. 91, No. 434. (Jun.), pp. 473-489.

Sanchez Munoz, 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
http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2011/48.e.pdf

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

Särndal, C.-E., B. Swensson, and J. Wretman. *Model Assisted Survey Sampling*, Springer Series in Statistics, 1992.

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

Skinner, C.J. and Shlomo, N. (2008) "Assessing identification risk survey microdata using log-linear models", *Journal Of The American Statistical Association* Volume: 103 Issue: 483.

Slacalek, J. (2009), "What Drives Personal Consumption? The Role of Housing and Financial Wealth", *The B.E. Journal of Macroeconomics*, Berkeley Electronic Press, vol. 9(1), pages 37..

Törmälehto, Veli-Matti, Olli Kannas and Markku Säylä (2012): "Integrated Measurement of Household-Level Income, Wealth and Non-Monetary Well-Being in Finland." Paper presented at the 32nd General Conference of The International Association for Research in Income and Wealth, Boston, MA.

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

UN (2005): "Designing Household Survey Samples: Practical Guidelines". United Nations Statistics Division, Studies in Methods Series F No. 98.

UN (2008) "Principles and Recommendations for Population and Housing Censuses (Revision 2)." *United Nations Statistics Division, Statistical papers Series* M No. 67.

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, Vijay and Gianni Betti (2008): "Data Accuracy in EU-SILC". In Atkinson, Anthony B. and Eric Marlier (ed.): *Income and Living Conditions in Europe*. Eurostat.

# APPENDIX: LIST OF HFCN PARTICIPANTS

| | |
|---|---|
| **National Bank of Belgium** | Philip Du Caju |
| | Isabelle Degreef |
| | Pierrick Stinglhamber |
| | Laurent Van Belle |
| **Danmarks Nationalbank** | Sigrid Alexandra Koob |
| **Deutsche Bundesbank** | Martin Eisele |
| | Julia Le Blanc |
| | Tobias Schmidt |
| | Ulf von Kalckreuth |
| | Junyi Zhu |
| **Eesti Pank** | Jaanika Meriküll |
| | Tairi Room |
| **Statistics Estonia** | Julia Aru |
| | Erika Taidre |
| **Central Bank of Ireland** | Mary Keeney |
| | Martina Lawless |
| | Yvonne McCarthy |
| | Aisling Menton |
| **Central Statistics Office Ireland** | Paul Crowley |
| **Bank of Greece** | Calliope Akantziliotou |
| | Antonios Loumiotis |
| | Theodoros Mitrakos |
| | Panagiota Tzamourani |
| **Banco de España** | Olympia Bover |
| | Ernesto Villanueva |
| **Banque de France** | Luc Arrondel |
| | Marc Chazelas |
| | Jérôme Coffinet |
| | Muriel Roger |
| | Frédérique Savignac |
| **INSEE** | Hélène Chaput |
| | Cédric Houdré |
| | Pierre Lamarche |
| **Bank of England** | Jeanne Le Roux |
| **Banca d'Italia** | Claudia Biancotti |
| | Andrea Brandolini |
| | Ivan Faiella |
| | Romina Gambacorta |
| | Stefano Iezzi |
| | Giuseppe Ilardi |
| | Andrea Neri |
| | Eliana Viviano |
| **Central Bank of Cyprus** | Ioanna Evangelou |
| | Stephan Haroutunian |
| **Banque centrale du Luxembourg** | Thomas Mathä |
| | Alessandro Porpiglia |
| | Michael Ziegelmeyer |
| **CEPS/INSTEAD Luxembourg** | Eva Sierminska |
| **Central Bank of Malta** | Karen Caruana |
| | Christopher Pace |
| **De Nederlandsche Bank** | Wilko Bolt |
| | Henriette Prast |
| | Federica Teppa |
| | Maarten van Rooij |

| | |
|---|---|
| | Corry van Renselaar[†] |
| **Oesterreichische Nationalbank** | Pirmin Fessler |
| | Martin Schürz |
| **National Bank of Poland** | Zbigniew Zolkiewski |
| **Banco de Portugal** | Sonia Costa |
| | Luisa Farinha |
| **INE – Statistics Portugal** | Carlos Marcelo |
| | Rita Rasquilha |
| **Banka Slovenije** | Uros Herman |
| | Matjaz Jeran |
| | Irena Komprej |
| | Vesna Lukovic |
| | Petra Ziherl |
| **Národná banka Slovenska** | Matúš Senaj |
| | Tibor Zavadil |
| **Suomen Pankki** | Risto Herrala |
| | Helinä Laakkonen |
| | Petri Mäki-Fränti |
| | Jouko Vilmunen |
| **Statistics Finland** | Olli Kannas |
| | Veli-Matti Törmälehto |
| **Eurostat** | Emilio Di Meglio |
| | Fabienne Montaigne |
| | Jean-Marc Museux |
| **European Central Bank** | Miguel Ampudia |
| | Katarzyna Bańkowska |
| | Frank Betz |
| | Robert Beyer |
| | Riccardo Bonci |
| | Dimitris Christelis |
| | Jeanne Commault |
| | Michael Ehrmann |
| | Ioannis Ganoulis |
| | Dimitris Georgarakos |
| | Reint Gropp |
| | Juha Honkkila |
| | Malgorzata Osiewicz |
| | Elena Pavlova |
| | Sébastien Pérez-Duarte |
| | Vitaliana Rondonotti |
| | Carlos Sánchez Muñoz |
| | Jiri Slacalek |
| | Frank Smets |
| | Philip Vermeulen |
| | Caroline Willeke |
| | Martin Zeleny |
| **Goethe University Frankfurt** | Michael Haliassos |
| **Einaudi Institute for Economics and Finance, Rome** | Luigi Guiso |
| **University of Naples Federico II** | Tullio Jappelli |
| **Board of Governors, Federal Reserve System** | Arthur Kennickell |
| **University of Oxford** | Peter Tufano |

[†] Corry van Renselaar passed away in January 2010.

# ERRATA

2 May, 2013

The following was corrected: in footnote 42, p. 95 the following sentence “For instance, in Spain, the Financial Accounts estimate of household debt in the last quarter of 2008 was 15% lower than the correspondent figure from banks’ balance sheets.” was corrected as “For instance, in Spain, the Financial Accounts estimate of household debt in the last quarter of 2008 was 15% higher than the correspondent figure from banks’ balance sheets.”