ANNEX A

Eurosystem household finance and consumption survey

In 2008, the Governing Council of the European Central Bank[1] decided to conduct a household finance and consumption survey (HFCS) in all euro area countries. The HFCS provides the Eurosystem with micro-level data on euro area households' wealth and consumption expenditure.

Why are central banks interested in micro information on household wealth and consumption?

In explaining the background behind this decision, it is important to note that there are policy questions for which the answer substantially depends on knowledge about distributional aspects that can only be assessed with micro-level data. For instance, the implications of steep increases in household indebtedness cannot be adequately judged from aggregate data alone. Information on the debt levels of detailed categories of households as well as their distribution across income and/or wealth classes obtained from surveys helps central banks to assess whether/to what extent the accumulation of household debt should give rise to concerns about households' financial soundness and, hence, may entail financial stability risks and/or other macroeconomic developments.

Due to the structural nature and low frequency of household finance and consumption surveys, their data are not meant to provide a direct input into high-frequency, day-to-day policy-making. However, results from research using the survey data inform policy-making and improve knowledge about economic structures and institutions, thereby providing important input into central bank policies. Looking beyond macroeconomic aggregates, information about the distribution of wealth, debt and income is important for better understanding the implications and transmission of macroeconomic shocks. Consequently, the HFCS data is a valuable input into a number of policy areas of the Eurosystem, *inter alia* monetary policy, macro-economic analysis and financial stability.

Prior to the decision to set up the HFCS, it was not possible to make an analysis of household wealth throughout the euro area because no household wealth surveys were conducted there. Furthermore, the surveys that existed in a few countries did not provide all the information needed on household assets and liabilities, and/or survey data were collected using largely incomparable methodologies. Given the large cross-country heterogeneity within the euro area in a number of relevant aspects (such as the importance of housing wealth, mortgage debt, etc., in household balance sheets), data from one country could not be used to infer the situation of others, thus impeding the possibility of reaching conclusions at the euro area level.

Consequently, in building up the foundations for the HFCS, one of the main priorities was the development and adoption of an *ex ante* agreed common methodology which would make it possible to undertake cross-country analyses based on comparable survey data as well as to calculate consistent indicators for the euro area.

General overview of the HFCS

The HFCS is a decentralised effort in which each participating institution (national central banks and, in a few countries, national statistical institutes) finances and conducts its own wealth survey.

The HFCS covers the whole euro area, with samples that provide representativeness both at the euro area aggregate level and at the individual country level.[2] This allows comparing the economic structure and the impact of different institutional features (e.g. banking system, pension schemes, household composition) on the transmission of economic shocks across individual countries.

Some countries conduct the survey every 2 years and others every 3 years.[3] The total euro area sample size is around 54 000 completed interviews (while country sample sizes vary from country to country, not necessarily in proportion to the country size). A longitudinal/panel component is available or planned in at least six country surveys.

The HFCS provides complete data sets for at least the basic components of household income, consumption and wealth. For information not collected due to item non-response, multiple-imputed values using correlations directly observed from the rest of the households are provided to users (see below). The significant effort applied in *ex post* fieldwork editing and imputation procedures entails a lapse between the end of the fieldwork and the time at which the data can be put at the disposal of final users.

For further details about specific features of the country surveys participating in the HFCS, see Table A.1.

Survey contents

General description

The Eurosystem HFCS questionnaire consists of two main parts: the first composed of questions referring to the household as a whole (answered only by the main respondent, namely the most financially knowledgeable household member) and the second targeted to individual household members (and answered by every household member aged 16 and over). The block covering household-level questions encompasses: real assets and their financing; liabilities and credit constraints; private businesses and financial assets; intergenerational transfers and gifts; and consumption/savings. Questions to individuals cover the following areas: demographics; employment, future pension entitlements; and labour-related income (other income sources being covered at the household level).

Given the focus of the survey on household wealth, priority is given to a detailed and accurate collection of survey information on household assets and liabilities. The objective of keeping the burden on respondents within reasonable limits implies that information on other items (such as on income or consumption) cannot be collected at the level of detail as in stand-alone surveys exclusively focusing on these themes. Nonetheless, the income information collected in the HFCS covers all household income sources, while information on consumption focuses mostly on specific recall questions which, according to the

Table A.1. **Main features of country surveys participating in the Eurosystem Household Finance and Consumption Survey**

	Responsible institution	Reference year	Frequency (years)	Panel component	Oversampling of wealthy households/criterion
Austria	Oesterreichische Nationalbank	2010/2011[1]	Three	No	No
Belgium	Banque Nationale de Belgique	2010	Three	Yes	Yes/income
Cyprus[12, 13]	Central Bank of Cyprus	2010[2]	Three	No	Yes/electricity bills
Estonia	Bank of Estonia	2013[3]	Three	Tbd	Tbd
Finland	Statistics Finland	2010[4]	Three	No	Yes[11]
France	Insee	2009/2010[5]	Tbd	No	Yes/taxable wealth
Germany	Deutsche Bundesbank	2010/2011	Two	Yes	Yes/certain areas
Greece	Bank of Greece	2009[6]	Three	No	Yes/certain areas
Ireland	Central Bank of Ireland	2011	Three	Tbd	Tbd
Italy	Banca d'Italia	2010[7]	Two	Yes	Yes/banking wealth
Luxembourg	Banque centrale du Luxembourg	2010/2011	Three	Yes	Yes/income
Malta	Central Bank of Malta	2010/2011	Three	No	No
Netherlands	De Nederlandsche Bank	2010[8]	Three	Yes	No
Portugal	INE Portugal/Banco de Portugal	2010[9]	Three	Tbd	Yes/certain areas
Slovenia	Banka Slovenije	2010	Three	No	No
Slovakia	Národná banka Slovenska	2010	Three	No	Yes/certain areas
Spain	Banco de España	2008/2009[10]	Three	Yes	Yes/taxable wealth

1. Data also available for the Survey on Household Financial Wealth in 2004 and the Household Survey on Housing Wealth in 2008.
2. Data also available for the Cyprus Survey of Consumer Finances in 1999 and 2002.
3. Since Estonia joined the euro area only in 2011, the Bank of Estonia will participate in the HFCS only as of the second wave.
4. Data also available for the Saving and Indebtedness Survey in 1987 and 1988, the Wealth Survey in 1994 and 1998 and the Housing and Wealth Survey in 2004.
5. High-income employees, self-employed and farmers.
6. Data also available for the Enquête Patrimoine in 1986, 1992, 1998 and 2004.
7. Data also available (not publicly though) for the Bank of Greece *Household Indebtedness Survey* in 2002 and 2005.
8. Data also available for the Survey on Household Income and Wealth starting in 1977 (and every two years since).
9. Annual panel data also available for the DNB Household Survey since 1993.
10. Data also available for the *Inquérito ao Património e Endividamento da Famílias: 2006* (for which cross-sections of 2000 and 1994 are also available).
11. High-income employees, self-employed and farmers.
12. *Footnote by Turkey:* The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus" issue.
13. *Footnote by all the European Union member states of the OECD and the European Commission:* The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

literature, provide information needed to accurately impute non-durable consumption in general purpose surveys (Browning, Crossley and Weber, 2002).[4]

When designing the HFCS contents, a large number of consistency checks were carried out between wealth and income survey variables and national accounts definitions. Although a large degree of consistency with national accounts definitions was a clear objective, some approximation is inevitable, given that survey categories must be translated into terms understandable to a broad population of respondent households.

Contents

The next two boxes summarise the contents of the HFCS (Figure A1) and the structure of the Eurosystem blueprint questionnaire (Figure A2).

Figure A1. **Contents of the HFCS**

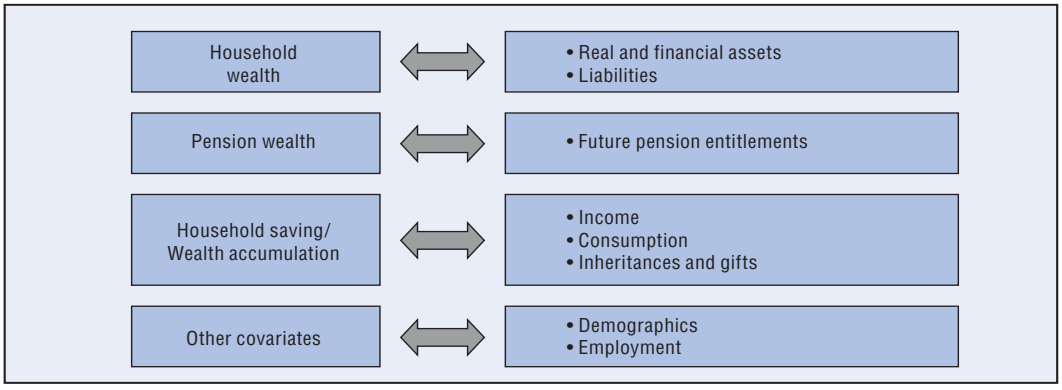
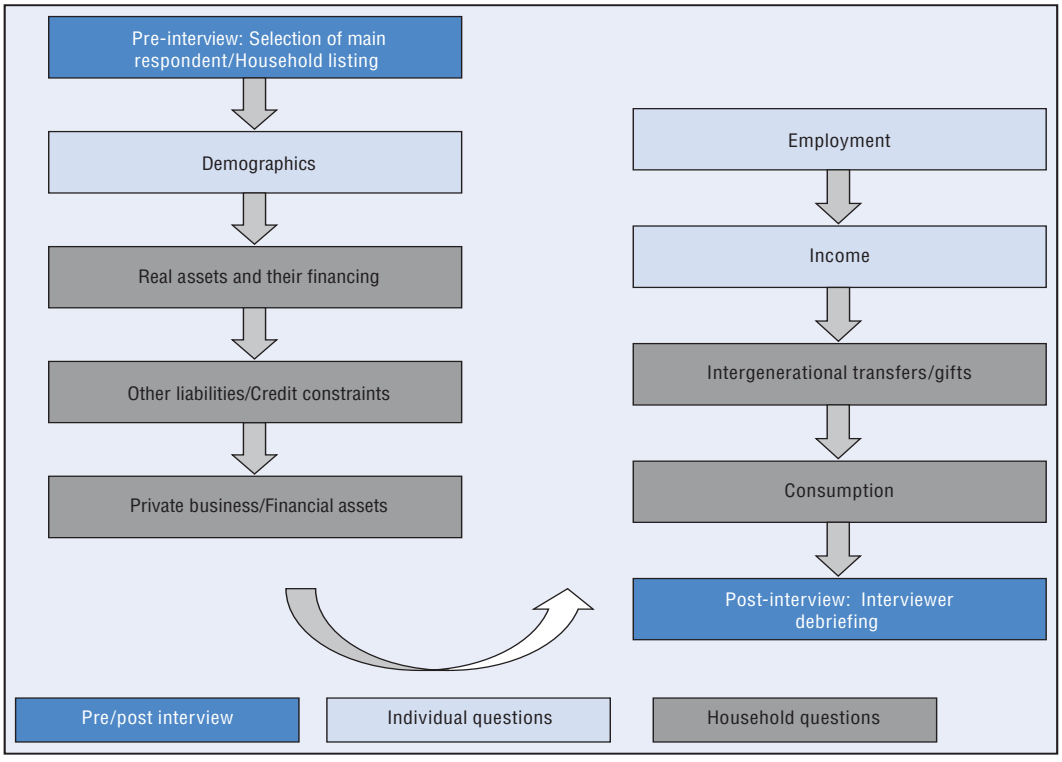


Figure A2. **The HFCS Blueprint questionnaire**



Methodological features of the HFCS

Ex ante *comparability through an output-oriented approach*

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. For this reason, the HFCS follows an output-oriented approach, i.e. all participating countries report core output variables based on a common set of *ex ante*-agreed definitions and descriptive features.[5] While questions in country surveys may be adapted to the specific circumstances, financial markets and products of each country, a common Eurosystem blueprint questionnaire is the starting

point for all country questionnaires. Some examples of the type of information collected within each category of the HFCS questionnaire are provided below:

- Demographics: age, gender, education, country of birth, marital status, relationship in household, etc.
- Real assets and their financing: tenure status; rent; different features of the household main residence (size, current value, year and price of acquisition, home equity withdrawal, etc.); features of other properties; vehicles; valuables; etc. Features of loans/mortgages collateralised by real assets: collateral, purpose, outstanding balance; maturity, monthly payments, year when loan was taken or refinanced; initial amount borrowed; loan refinancing, fixed/adjustable interest rates, etc.
- Other liabilities/financial constrains: overdrafts; credit lines; credit-card borrowing; leases; consumer/instalment loans; etc. For the most important loans: purpose, outstanding balance; maturity; monthly payments; initial length; initial amount borrowed; etc. For remaining loans: outstanding balance and monthly payments.
- Private businesses: self-employment businesses; passive investments; activity; legal form; employees; value; etc.
- Financial assets: sight accounts; savings accounts; mutual funds; bonds; publicly traded shares; shares in foreign companies; managed accounts; informal loans to relatives or friends; investment attitudes; etc.
- Employment: employment status; main employment; employment history; expected age of retirement; etc.
- Future pension entitlements:[6] features of government-sponsored, occupational and private pension schemes, life insurance, etc.
- Income: 12-month gross income by individual sources; comparison with average; next-year expectations; etc.
- Intergenerational transfers/gifts: for the most important ones: when they were received; how much; from whom; expected inheritance; etc.
- Consumption: food in and out of the home; regular household transfers (alimony, assistance, etc.); comparison with average and with income; saving motives; emergency assistance.

Specific features of the common HFCS methodology

Probability sampling

All HFCS country surveys should have a probabilistic sample design, i.e. each household in the target population should have a non-zero probability of being selected and that probability should be known *ex ante* for all households.

Oversampling of the wealthy

Representing the behaviour of the typical (average) individual household can be achieved through standard proportional sampling methods. However, a purely random selection of units would yield a statistically inefficient estimate of the distribution of wealth. On the one hand, wealth is highly unevenly distributed, i.e. a small share of households holds a disproportionately high share of total wealth. On the other hand, portfolio sophistication also increases with wealth, such that certain financial instruments are almost exclusively held (and in large quantities) by the wealthiest households. If such

households are not well represented in the final sample the results may likely not provide a reliable picture of the distribution and composition of household portfolios. A given level of precision would either require a rather large (and costly) sample or, if efficiently designed, a sample that should include a higher-than-proportional number of wealthy households (Muñoz, 2011).

Achieving an adequate portion of wealthy respondents is challenging: first, establishing contact with wealthy respondents may be more difficult as they are usually surrounded by additional security measures, may be absent from their principal residence during prolonged periods of time and usually possess more than one residence. Second, both available time as well as the self-perceived value/time ratio often pre-dispose wealthy households to refuse more frequently to participate in surveys.

Consequently, representing the total mass of wealth requires sophisticated sample designs and contact strategies. All in all, while oversampling the wealthy may add to total survey costs, it increases precision and reduces non-response bias. In addition, oversampling also improves efficiency in the estimation of variables positively correlated with wealth. For that reason, wealthy households are oversampled in most euro area countries, using country-specific techniques (see details in Table A.1).

Survey mode

Survey information is collected through Computer-Assisted Personal Interviews (CAPI),[7] i.e. face-to-face interviews administered by an interviewer using a computer to record the replies provided by respondents. The use of a computer allows a smooth and error-free administration of the routing of the questions (which is quite 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 terms of: *i)* persuading respondents to participate in the survey/increasing response rates and reducing the risk of response bias; *ii)* building up trust *vis-à-vis* respondents, thus lowering the likelihood that a respondent drops out in the middle of the interview; *iii)* minimising levels of item non-response by personally assisting (i.e. offering pre-designed prompts) – if required – during the interview; *iv)* avoiding incomplete responses; and *v)* providing additional information (interviewers' observations/*paradata*); etc.

Multiple imputation

Imputation assigns a value to a variable when it was not collected or not correctly collected. Standard econometric tools can only deal with complete data sets. Consequently, it is difficult (although possible) to use the data without imputing missing values. Leaving the imputation to the users of survey data is one option, but this is dangerous in terms of quality of the results, ease of use of the data, and potential misuse. Because data producers have access to confidential information about the reasons for non-response and other information not released for public use, they are in a better position to impute than general users of the data. At the same time, as long as imputed variables are appropriately flagged, it remains the users' choice to work with imputed or not imputed data.

The HFCS provides multiply-imputed values to cover for item non-response for at least the basic components of household income, consumption and wealth. To that end, stochastic imputation is applied, meaning estimating missing observations conditional upon observed variables that can plausibly explain missingness. For each missing value, five imputed values are estimated (thus giving rise to the same number of complete data sets). The reason why missing values are multiply imputed is that if the procedure were run just once (single imputation), without adding the appropriate random term, it could yield information that does not take into account uncertainty (the resulting variance could be underestimated). This would be a particular problem in cases of significant item non-response.

Construction of survey weights

To improve the quality and comparability of the analysis, it is essential that initial sample weights are adjusted to compensate for various features and/or imperfections in the sampling design and in the final sample. The standard procedure for computing and adjusting survey weights entails weighting factors to take into account: *i)* the unit's probability of selection (design weights); *ii)* unit non-response (non-response-adjusted weights); and *iii)* adjustment of the weights to external data (calibration, post-stratification, etc.) to approximate the sample to the distribution of households and persons in the target population (final weights).

For the HFCS, weights are calculated based on the following sequence: design weights – coverage adjustment – non-response adjustment; adjustment to external sources and replicate weights. As mentioned, HFCS design weights are calculated for all units selected in the initial sample as the inverse of the selection probability of each unit (probability-sampling).

As for coverage adjustments, the definition of the target population is adjusted in cases where some groups of households cannot be covered by the sampling frame (instead of introducing adjustments to the weights). Weights are also adjusted for over-coverage, multiple-selection probabilities, for non-response (via estimated response probabilities of homogenous response groups).

Weights are then adjusted to external data sources with calibration to margins to match the corresponding population totals and category frequencies. The choice of calibration variables should be such that control variables are strictly comparable to the corresponding survey variables, correlated with the study variables, but not too correlated with each other. Although the selection of calibration variables is country-specific, the weighted distribution of the sample by gender, age and household size should be equal (or close) to the corresponding distributions in the population. The sum of the final weights in the sample must also be equal to the number of households in the population.

After the calculation of final weights, replicate weights should be calculated to estimate the variance of the estimates (see below). Replicate weights are adjusted for non-response and calibrated according to the same procedures used to adjust final weights.

Variance estimation

Variance estimation is an essential element in the use 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 an estimator can only be known with certainty if the whole population is observed. Underestimating the variance of an estimate may lead to incorrect conclusions,

while overestimating the variance decreases the usefulness of the data, as fewer outcomes are statistically significant. Complex sampling designs require variance estimation procedures that are more complex than the standard ones utilised in surveys with simple random samples from a large population.

In order to allow users to estimate variance for the HFCS, countries provide 1 000[8] replicate weights using a bootstrap replication method.[9] 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 be used as well in the case of multi-stage samples drawn with a low sampling fraction in the first stage.[10] While – as with 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), the method provides consistent variance estimates in case of non-smooth statistics such as quantiles.

Since the final weights are adjusted for non-response, post-stratified or calibrated (the specific technique not being important), the replicate weights are adjusted as well according to the same procedure (e.g. by running CALMAR with the same margins on each of the replicate weights). This can be considered as an additional rescaling factor. For instance, after drawing the sample and rescaling the weights, the weights are further rescaled to satisfy post-stratification or calibration constraints for each replicate. This ensures that the replicate estimates are close to unbiased in each replicate sample.

Survey evaluation of the HFCS

Following the completion of each HFCS wave, a comprehensive evaluation of the survey will be carried out. This evaluation shall encompass an exhaustive quality assessment of the survey results (with special focus on cross-country comparability) summarised in a quality report intended for public release. A review of user requests and a critical evaluation of the extent to which survey questions meet the HFCS objectives will also be conducted; such a review should lead to a regular update of the common blueprint Eurosystem questionnaire with a view to subsequent waves of the HFCS.

Intended outputs

The first dissemination of the HFCS research data set is planned for early 2013. Anonymised micro data will be made available to the research community. Access to the data will be granted upon successful evaluation of individual research proposals guaranteeing non-commercial use of the data and also on the condition that researchers fulfil safety conditions on data storage, sign a confidentiality commitment, etc.

In addition, a set of euro area aggregate indicators (on portfolio and debt composition; the debt burden of indebted households; saving, consumption and access to finance; distribution of wealth; etc.) will be released alongside the HFCS micro data set.

Further information on the HFCS can be obtained on the HFCS website: *www.ecb.europa.eu/home/html/researcher_hfcn.en.html*.

Notes

1. The ECB Governing Council is the body governing the Eurosystem, i.e. the ECB and the National Central Banks of the 17 countries that have adopted the euro.

2. 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.
3. Some countries chose a higher two-year frequency because their surveys have a panel component (meaning that the households participating in the panel are interviewed again in subsequent waves). Every three years is deemed too low a frequency to keep track of and properly follow panel households.
4. Browning, M., T.F. Crossley and G. Weber (2002), "Asking Consumption Questions in General Purpose Surveys", *SEDAP Research Paper,* No. 77.
5. In addition, a set of standardised non-core extensions are included in some (but not all) country questionnaires.
6. For current pension entitlements, in the following section information is collected on the amount of pension income collected in the last 12 months.
7. The only exception is the Netherlands (survey conducted by the Nederlandsche Bank), where a CAWI mode (computer assisted web-based interviewing) is in place.
8. 1 000 is a commonly used compromise between computational efficiency and stability of the variance estimates.
9. Bootstrap was selected over other methods (e.g. Jack-knife or balance repeated replication) because it allows analysts to select the number of replicates (in other methods the number of replicates is determined by the number of strata and/or number of primary sampling units, PSUs). Besides, bootstrap samples are independently drawn across strata, so the replicate weights of different countries can be stacked and analysed as if they came from a single bootstrap procedure, thus allowing users to calculate a variance in the combined euro area data set in a standard way.
10. This is the case in several popular setups of stratified sampling. In addition, other sampling designs can be approximated by this setup.