Statistics Paper Series

Pierre Lamarche

Estimating consumption in the HFCS

Experimental results on the first wave of the HFCS

No 22 / May 2017

Contents

Abstract

In this paper, we estimate consumption in the first wave of the Eurosystem Household Finance and Consumption Survey for a subset of countries that account for around 85% of the aggregate final consumption expenditure of households in the euro area. For this purpose we use the methodology described by Browning et al. (2003), taking advantage of the few questions on consumption asked to households participating in the survey and information on consumption collected in the Household Budget Surveys. Using also the framework developed for statistical matching, we give assessments of the uncertainty related to this kind of estimation. We find that the quality of estimation varies greatly across countries and in general is sensitive to the Conditional Independence Assumption implicitly made through this exercise. At any rate, estimations of consumption (provided throughout this paper) should be used with caution, bearing in mind that they rely on strong assumptions.

Non-technical summary

When looking at macro-data, households' consumption expenditure turns out to be an essential component of GDP, evolving in a parallel way over time and contributing significantly to it. Explaining and understanding consumption behaviours is therefore essential with regard to economic policies, not least because of the long-term consequences of saving behaviours on capital accumulation. Across euro area countries, levels of mean consumption have proven to be quite heterogeneous and, depending on the countries, to have reacted significantly to the crisis in 2008.

Some facts are already well documented regarding consumption behaviours and may be observed in the data: thus, average consumption varies with age, and in particular it tends to drop after 60. It naturally increases with income and moreover, the share of food in consumption expenditure is expected to decrease with income. This last pattern is however not verified for some of the countries of the euro area. Finally, savings, defined as the residual between income and consumption, follow a pattern consistent with the life cycle theory: strongly negative at the beginning of active life, it increases as time passes, then tends to decrease with retirement and the subsequent drop in income. Finally, the higher current income is, the more households save as a share of their income.

Our goal in this paper is to link households' income, consumption and also wealth. For this purpose we use the data from the Household Finance and Consumption Survey (HFCS); this survey coordinated by the European Central Bank, is aimed at collecting at household level detailed information on assets and liabilities. We have also at our disposal information on gross income and a few variables related to some sub-items of consumption (mainly food consumption). Using another European survey on consumption, the Household Budget Survey (HBS), we are able to link, through modelling, a given level of consumption expenditure, along with some variables describing the household, with a level of consumption for non-durable goods and services. We apply this link to a sample of countries. As a result, we face a dataset containing a representative sample of households; for each of these households, we have at our disposal data on income, wealth and consumption. The imputation reproduces relatively well the expected distribution and links with other variables, although the quality of the imputation may vary quite significantly across countries.

Such an exercise is based on strong hypotheses and in particular, it assumes that once the level of food consumption and the demographic variables are taken into account, the level of consumption is independent from another variable that we would like to include in the analysis – wealth for instance. We implement different methods in order to estimate the uncertainty related to the estimates we may develop from such an exercise. The most basic methods show that the estimations remain quite stable. Nevertheless, using experimental methods to relax our main assumption of conditional independence assumption leads us to far wider intervals of plausibility, thereby proving that the estimation relies a lot on this assumption. To

conclude, in case such an assumption turns out to be wrong, this would jeopardise the entire estimate. Therefore, conclusions drawn from these data have to be made cautiously.

1 Introduction and motivation for estimating consumption in the HFCS

1.1 Consumption as a key element of GDP

Understanding households' consumption is essential in economic analysis as consumption expenditure accounts for an essential part of national GDP in most developed countries. In particular, between 2000 and 2013, households' total expenditure typically accounted for 56% of GDP, a very steady share across the years.

Chart 1

GDP and final consumption expenditure per capita in euro area

(current EUR in thousands; the definition of the euro area is kept constant over time, encompassing 18 countries as of 1st January 2014)
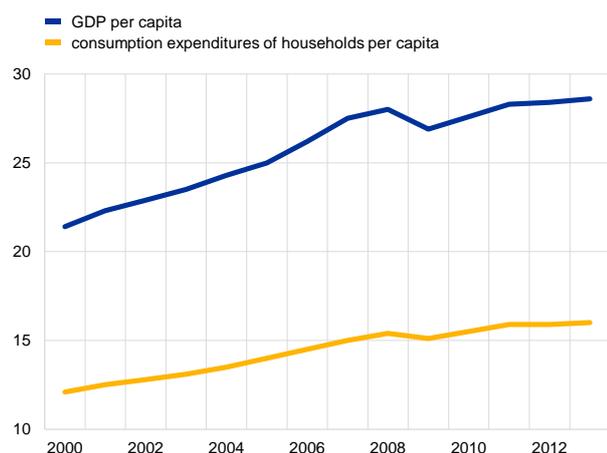
Sources: Eurostat.

As shown on **Chart 1**, GDP per capita reached EUR 28,600 in 2013, while households' consumption expenditure per capita was up to EUR 16,000. The two aggregates have evolved in a similar manner over the period, even if consumption turned out to be less sensitive to the crisis in 2008 than GDP.

This is confirmed by the analysis at national level of the importance of households' consumption as a proportion of GDP. Indeed, the share of consumption grew in most of the countries belonging to the euro in 2008 and 2009, reflecting the resistance of households' consumption to negative shocks (**Chart 2**).

Chart 2 reveals also a strong heterogeneity across euro area countries with regards to the importance of households' consumption as a proportion of GDP: thus households' consumption accounts for about 72% of Greek GDP in 2012, while it represents only 30% of the GDP for Luxembourg. Most interestingly, the share of consumption decreased strongly over the period, by more than 10 points. Such a decreasing pattern may be observed only for two other countries, Malta (-7 points between 2000 and 2013) and the Netherlands (-4 points).

Chart 2

Households' consumption expenditure as a share of GDP
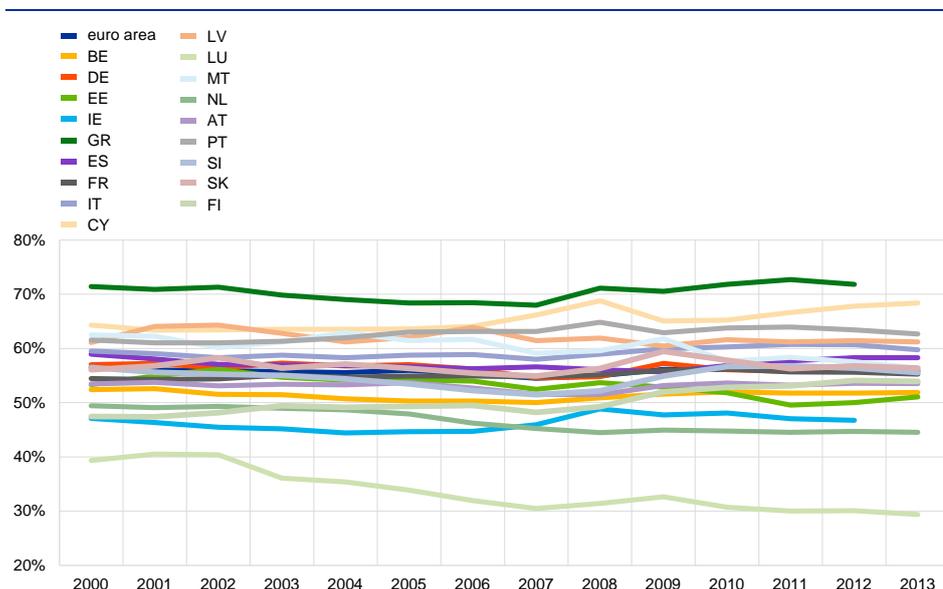
Sources: Eurostat.

As shown on **Chart 3**, consumption plays an important role in GDP growth, which is of course the direct consequence of its importance in total GDP. Between 2000 and 2008, consumption contributed to between 1 and 2 points of GDP growth, which represents a strong contribution given the growth rate of GDP over this period. The consequence of more steady consumption from 2008 onwards, its contribution has turned out to be weaker in recent years. However, the fact that the growth of GDP has also slowed tends to support the idea that consumption is an essential determinant of GDP growth.

Chart 3

Contribution to final consumption expenditure to euro area GDP growth (in % point)
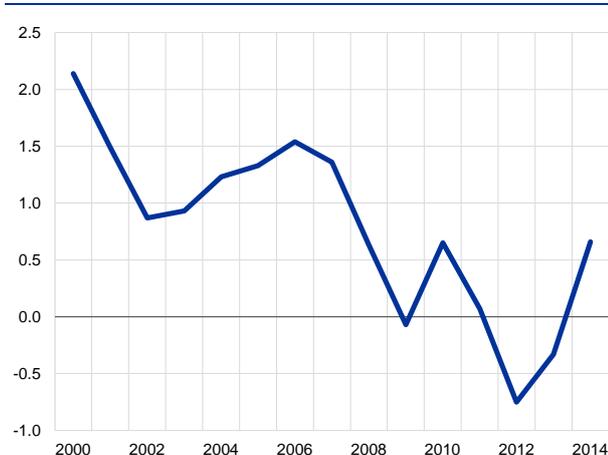
Sources: Eurostat.

Finally, consumption represents the ultimate goal of all economic activity. Indeed, one of the main assumptions of economic theory consists in the fact that the more individuals consume, the more their well-being increases. From this point of view, describing and understanding the link between consumption and wealth or income is essential.

1.2 Consumption is driven, at least partly, by monetary policies

As a result, the estimation of the impact of economic policies on consumption is essential for assessing the efficiency of such policies. A large body of literature aims to examine such effects (see for instance Lettau et al. (2002), Muellbauer (2010), Carroll et al. (2011) or Aron et al. (2012), essentially on U.S. data).

Turning now to Europe, estimations of wealth effects for some countries belonging to the euro area have been made by Slacalek (2009), following the same methodology as Carroll, et al. (2011). These estimates based on macro-data show significantly lower wealth effects in countries belonging to the euro area than in countries outside the euro area[1]. This paper also addresses the issue of the heterogeneity of wealth effects according to the type of assets; indeed, the reaction of households to shocks according to the liquidity of the assets which those shocks impact. The authors find a significantly lower if any marginal propensity to consume out of housing wealth rather than financial wealth (see **Table 1**).

Table 1
Marginal propensity to consume out of wealth according to euro area countries

(stars on the left side of the figure indicate the statistical significance of the estimations: * 10%, ** 5". *** 1%)

Country	Total wealth effect	Housing wealth effect	Financial wealth effect
AT	0.14	0.40	-2.17
BE	-0.02	0.63	-6.74
DE	3.32*	14.24	2.86
ES	2.38***	5.33**	6.24**
FI	4.03***	-3.58	18.15***
FR	1.07	2.89*	2.30
IE	1.84*	2.09	9.15*
IT	-0.33	10.30*	-1.07*
NL	1.14**	2.68*	3.17

Note: Marginal propensities to consume consist of the cumulated effects on consumption resulting from an initial shock to wealth.
Sources: Slacalek (2009).

Wealth effects turn out to be strongly heterogeneous according to the type of assets, but also the institutional conditions regarding housing and financial markets. At the same time, differences in household portfolios may explain the various reactions to an increase (or decrease) of an asset's price. Indeed such statistics are computed at the macro-level; in that sense, micro-data may help to better take into account heterogeneous effects of shocks experienced on the different types of assets.

---

[1] In this paper, countries belonging to the euro area are: AT, BE, ES, FI, FR, IE, IT, NL.

1.3 Understanding wealth accumulation is also key to the long-term perspective

As already mentioned, on the one hand, understanding consumption dynamics is essential for evaluating the effect of monetary policies on the economy. On the other hand, the description of wealth accumulation and its determinants is also of huge interest, as it ultimately will determine the distribution of a non-negligible part of income. From this point of view, the heterogeneity of saving rates plays a crucial role which has to be accurately described and understood.

According to National Accounts, the gross saving rate ranges between -7.6% of disposable income in CY and 16.6% in DE which highlights strong differences with regards to the saving rates across countries. Saving rates may be affected in the short term by unexpected increases or decreases in disposable income. Therefore asymmetric shocks to income that may affect the euro area may lead to an increase in the heterogeneity in saving rates. However, the range observed in the euro area remains somewhat constant over time (see **Chart 4**). This seems to reveal rather structural differences and national specificities in the way European households tend to accumulate wealth across countries.

Chart 4

Evolution of gross saving rate in the euro area between 1995 and 2013

(percentage points)
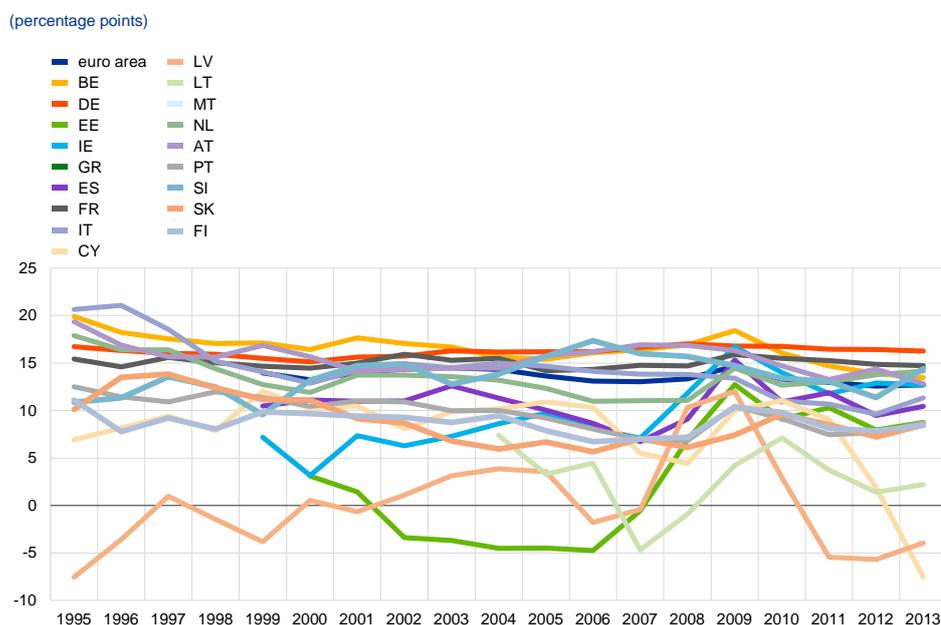
Sources: Eurostat.

The analysis of such differences may be performed using micro-data which could help to better account for the heterogeneity between households and explain the differences observed across countries. From this point of view, there is no source of information to measure precisely saving flows among households. Thus it is necessary to take advantage of already existing surveys in order to construct the required information on households' savings.

Also, special concerns have to be taken into account regarding the precise definition of the saving rate. As shown for instance by Audenis et al. (2002), the level of saving rate is highly sensitive to its definition. Needless to say, this definition should be the most homogeneous possible across countries. It affects not only consumption (through durables, transfers in kind), but also income (disposable income vs. adjusted disposable income, indirect taxes…).

1.4 A reinforced need to focus the economic analysis on households

Published in 2009, the Sen-Stiglitz-Fitoussi report underlines the need to focus more closely on the household sector, as it is more likely to reflect human well-being. In particular, it states that the emphasis should be put on consumption and income rather than GDP; it also insists on the importance of taking into account distributional indicators as average values usually provided by National Accounts do not give a full picture of the resources individuals have at their disposal in order to finance their needs. Finally, it call for a better description of the link between income, consumption and wealth, as these three dimensions are essential in order to understand how households allocate their resources, also in view of analysing financial sustainability for households.

From this point of view, surveys aimed at collecting information on households' balance sheet, income and consumption should be used in order to better assess the distributional aspects of these dimensions. There is still room for improvement of the statistical information system, as there are few variables on consumption in surveys such as HFCS and EU-SILC, and also very few details on assets, indebtedness or income in surveys such as HBS.

2 Consumption in the euro area: what do we already know?

2.1 Repartition of consumption across the euro area

Households in the euro area consumed an average yearly amount of EUR 36,400 in 2013; this figure was EUR 33,800 in 2005[2] (**Chart 5**). This figure hides strong differences between countries: average consumption was EUR 78,200 in LU in 2013, more than 4.5 times higher than in EE, LT or LV. One of the most noticeable evolutions is the one in IE: average yearly consumption reached EUR 54,700 in 2007, before strongly decreasing during the crisis, losing 16% of its value between 2008 and 2009. From this point of view, the evolution of consumption following the crisis took different paths depending on the countries.

Chart 5
Evolution of average household consumption in nominal terms in the euro area

(thousands in EUR)
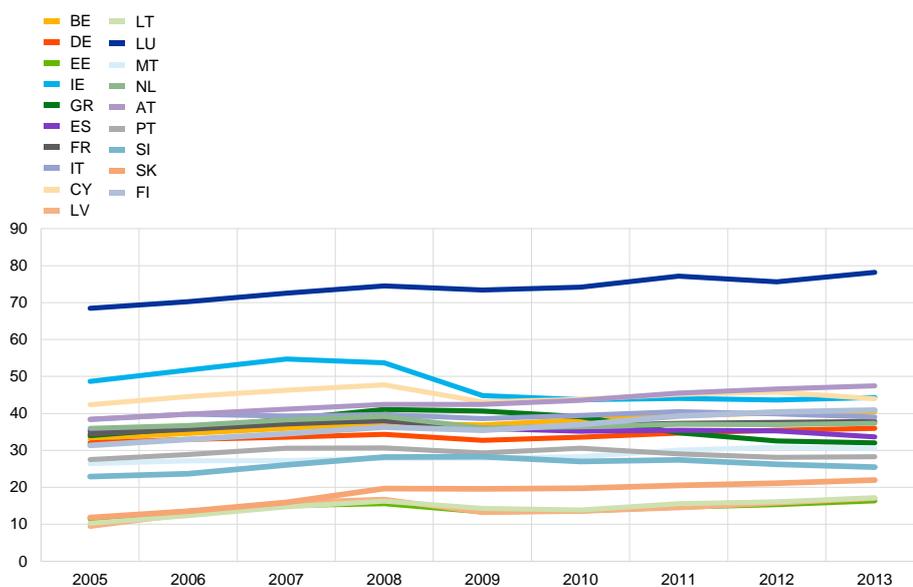
Sources: Eurostat.

The evolution of consumer prices is also heterogeneous across the euro area: the HICP increased by 18% between 2005 and 2014 in the euro area, but this increase ranges from 10% for IE to 40-50% for the Baltic States (see **Chart 6**).

---

[2]   The euro area is kept unchanged across time, encompassing 19 countries as of beginning of 2015.

Chart 6

Annual evolution of HICP in the euro area
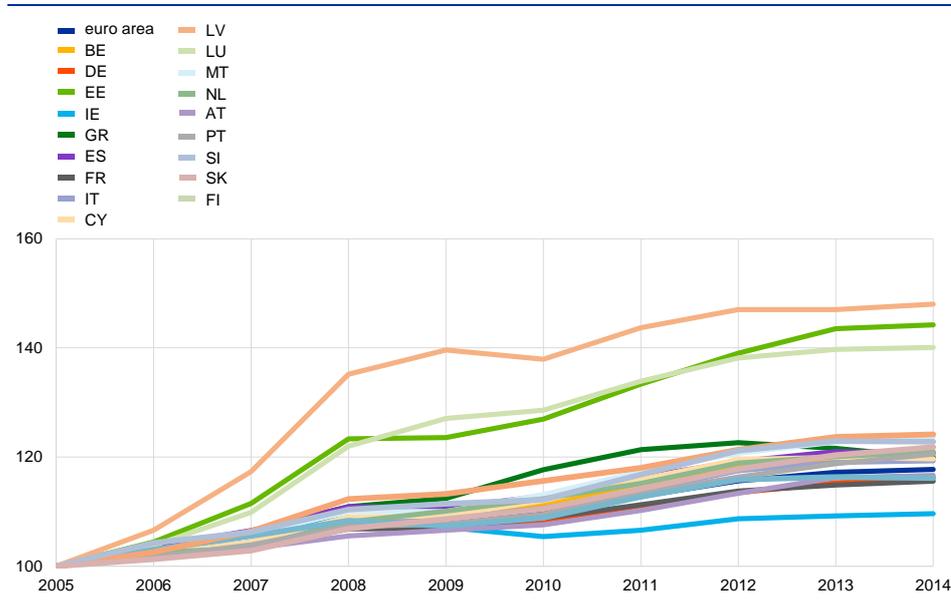
Sources: Eurostat.

Taking into account the evolution of consumer prices delivers a quite different picture of consumption in the euro area over the period: between 2005 and 2013, consumption in real terms decreased by 10%, with a turning point in 2009, when average consumption has decreased by EUR 2,200 (2005 prices). The most striking drop concerns IE, where consumption decreased by 37% in real terms over the period, and especially in 2009 with a decrease of EUR 7,200 on average. Conversely, SK or LV experienced a 50% increase of their average consumption.

When looking at the composition of consumption and its evolution over the period, it appears that the breakdown between durable goods, non-durable goods and services has remained quite stable: the share of durable goods in euro area's consumption went from 10% in 2005 to 8% in 2013, while services represented 51% of consumption in 2005 and 53% in 2010. As for the remaining elements, semi-durable goods represented 8% and non-durable goods accounted for 31% of total consumption in 2013. Consumption of durable goods ranges from 4% of total consumption in GR, CY or LV, to 11% in DE. The share of non-durable goods is very volatile: it ranges from 27% in AT to 54% in LV; the same applies approximately to services with a minimal share of 28% in LT and a maximal share of 58% in ES in 2013.

If we look at a more detailed breakdown of consumption, we observe that the composition of the average basket for households in the euro area is highly stable (**Chart 7**). Following the COICOP classification, the breakdown shows for instance that food consumption accounts for 12% of total consumption in the euro area. This figure may vary depending on the county: in the Baltic States, it ranges from 20 to 23%, while it represents only 10% of total expenditure in AT, DE, IE and LU. Another important item of consumption is spending on housing and certain utilities (water,

gas and fuel): its share was 24% in 2014. Here again, this share may differ strongly depending on the country: it ranges from 11% in MT to 28% in FI.

Evolution of households' consumption basket in the euro area between 2005 and 2014

- CP01 - food and non-alcoholic beverages
- CP02 - alcoholic beverages, tobacco and narcotics
- CP03 - clothing and footwear
- CP04 - housing, water, electricity, gas and other fuels
- CP05 - furnishings, household equipment and routine household maintenance
- CP06 - health
- CP07 - transport
- CP08 - communications
- CP09 - recreation and culture
- CP10 - education
- CP11 - restaurants and hotels
- CP12 - miscellaneous goods and services

Sources: Eurostat.

Likewise, transport expenses account for 13% of the total expenditure in the euro area, ranging from 7% in SK to 17% in LU.

The allocation of expenditure across the different items proves to be very stable over time, as shown by the evolution of the composition of the households' basket. Unexpectedly, the crisis in 2008 does not seem to have affected those decisions. More surprisingly, the prices for the different categories have evolved in very different ways, without causing any significant change in the structure of consumption. Between 2005 and 2014, the price index for food and non-alcoholic beverages rose by almost 18%, similar to the evolution of the global HICP over the period. At the same time, prices for alcoholic beverages, tobacco and narcotics increased by 36%, prices for housing and transport went up by 30% and 24% respectively, while recreation and clothing prices remained almost stable (+3% and +6% respectively). Telecommunication prices actually saw a strong decrease over the period, by 19%.

Neither the prices nor the evolution of income seem to have an effect (at least over a 10-year period) on the composition of households' expenditure. Of course, this statement is valid at the macro-level; nothing indicates that this is also true at the individual level. However, in case households modify the composition of their basket

according to the evolution of the different prices, then this reveals a strong heterogeneity of reactions across the population.

## 2.2   The Household Budget Survey: a first insight into understanding consumption

Aside from the information provided in National Accounts, we focus on information on consumption at the micro-level. Indeed, aggregate statistics provide reliable information on the evolution of average consumption in the euro area; however it does not reveal everything about the heterogeneity that may affect households along, say, the distribution of income or wealth. In order to better assess the choices made by households regarding consumption and savings, collecting information at the same time on wealth, income and consumption at the micro-level is essential.

The Household Finance and Consumption Survey (HFCS) conducted in the euro area by National Central Banks, is aimed at gathering information on these three particular aspects at the household level, with important differences in terms of details. Wealth and indebtedness are very well described in the questionnaire, while the survey only collects approximate pieces of information about income and actually contains only three or four questions on consumption, essentially focusing on food consumption. Therefore, the HFCS cannot pretend to shed light on households' consumption behaviours per se.

Another important source of information regarding consumption is the Household Budget Survey (HBS) which is conducted in all EU countries by the National Statistical Institutes. The primary aim of these surveys is the computation of weights for the Consumer Index Price. As such, these surveys collect very detailed pieces of information on households' expenditure through diaries that households participating in the survey are asked to fill in. Such a method turns out to be inapplicable in the case of the HFCS, the questionnaire for which is already very long and demanding. However, the HBS data already contain some insights into consumption behaviours in the euro area.

Chart 8

Coverage rate for total consumption in the HBS data as compared with National Accounts
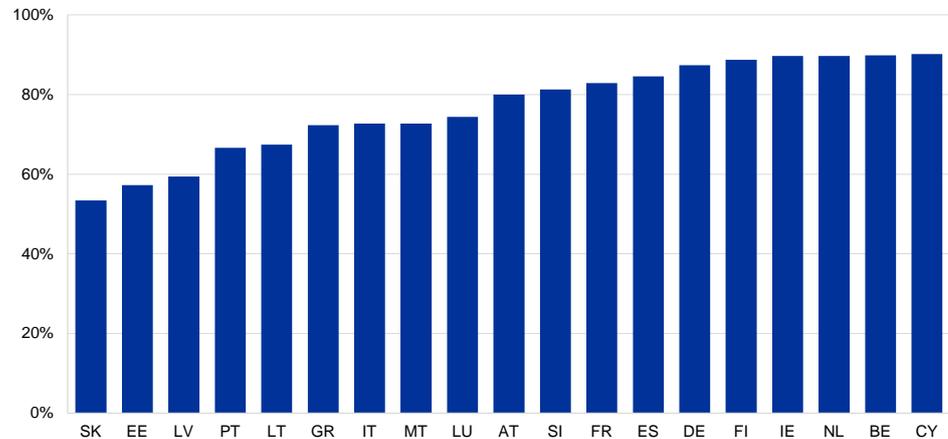
The HBS was conducted in the Member States of the European Union in 1988, 1995, 1999, 2005 and 2010. We will focus on the results of the last wave of HBS, as it incidentally coincides with the first wave of the HFCS.

First of all, average consumption as measured by the HBS is different from the one given by National Accounts. Indeed, according to HBS data, average consumption per household in the euro area in 2010 was EUR 29,200 and EUR 27,600 in 2005. Such figures account for 82% of the aggregate amounts available in the National Accounts; the evolution over time remains consistent between the two sources of information. The coverage also differs strongly across countries (see **Chart 8**): it is 90% in countries such as BE, IE, CY or NL, while it is close to 50% in SK. The differences between the National Accounts and the surveys are due, as is normally the case, to differences in the target population (the surveys do not collect information for households living in institutions), as well as conceptual differences. In particular, with the introduction of ESA2010, there may be a significant amount of imputed expenditure regarding consumption that comes under item 12 (Miscellaneous goods and services).

Chart 9

Ratios to the average consumption according to age of the reference person
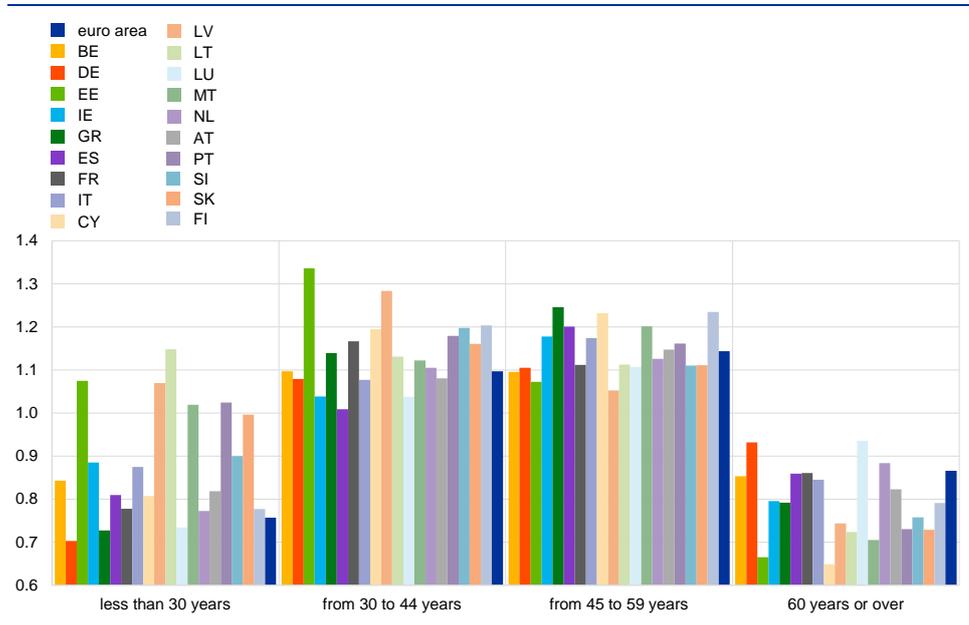
Sources: Eurostat.

Regarding the main results shown by the HBS data, age turns out to have a significant effect on the amount of expenditure. The amount of consumption expenditure follows a hump-shape curve: hence, households in the euro area whose reference person is below 30 spend 24% less in total consumption than the average; older people are in a similar situation with a total consumption 13% below the average. However, the situation is highly heterogeneous across countries, as shown in **Chart 9**, especially for the people under 30.

A far more important determinant for consumption is of course income. As shown in **Chart 10**, the variation across the income distribution is large and appears to explain more comprehensively the amount of expenditure. In particular, the variation across countries is less strong. On the one hand, households in the euro area and belonging to the first income quintile spend on average 52% less than the total population; on the other hand, households belonging to the last quintile consume in one year 73% more than the average.

Chart 10

Ratios to the average consumption according to the income quintile

■ euro area	■ LV		
■ BE	■ LT		
■ DE	■ LU		
■ EE	■ MT		
■ IE	■ AT		
■ GR	■ PT		
■ ES	■ SI		
■ FR	■ SK		
■ IT	■ FI		
■ CY			
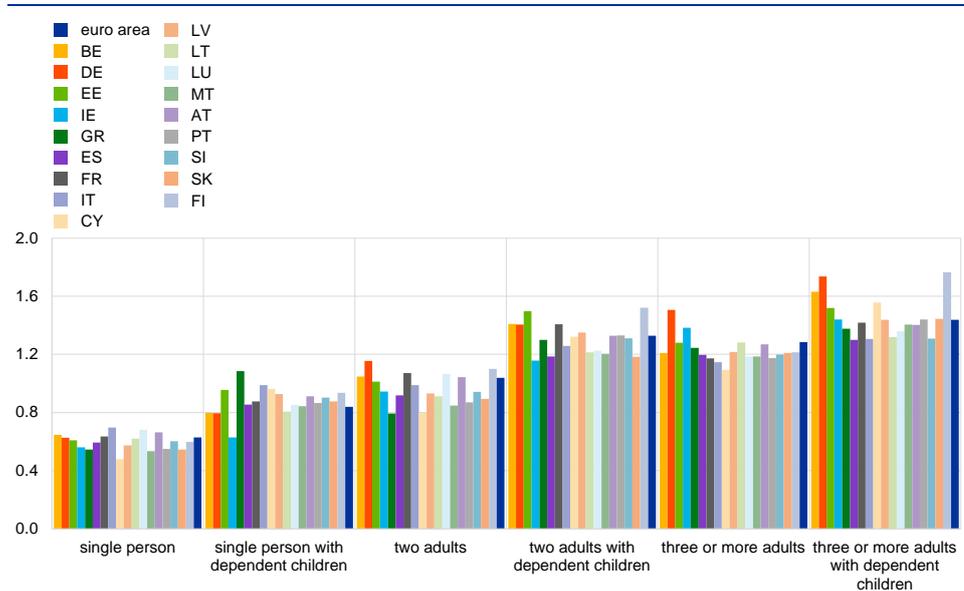
Other variables, such as the socio-economic category or the household structure, may help explain the amount of consumption for a given household. Thus, in 2010 households in the euro area where the reference person was unemployed spent 38% less on average than the total population; retired people also consumed on average 16% less. However, the effect of this particular variable has to be estimated for given income; from this point of view, micro-data are crucial as they enable micro-econometric analyses to be carried out.

Finally, as expected, the size of the household has an impact on the expenditure of the household. Single persons consume on average 37% less than the total population; households with three or more adults and dependent children spend 44% more. This last figure may even be higher in some countries, such as in FI, where it reaches 76%.

Another important part of the analysis has to be dedicated to the composition of consumption, and in particular to the share of food consumption. In particular, one major finding in consumption theory states that the share of food decreases as income rises[3]. This result holds for most of the euro area countries: 12 out of 19 countries exhibit such a pattern. However, as shown in **Chart 11**, this is not the case for some countries, for instance FR or NL, for which the evolution of the share of food follows rather a hump-shaped curve. This leads to the fact that overall, in the euro area, the Engel curve does not seem to hold, as the share of food for the first income quintile (14%) is lower than the proportions for the higher quintiles (respectively 16%, 15%, 15% and 14% for the second, third, fourth and last quintiles). However, this finding is probably valid only with income quintiles defined at

---

[3]   This result is known as the Engel's law.

the national level. Indeed, given the heterogeneity across the different countries, it is very likely that quintiles defined at the euro area level would verify the Engel curve assumption.

Chart 11

Share of food in total consumption according to income quintile
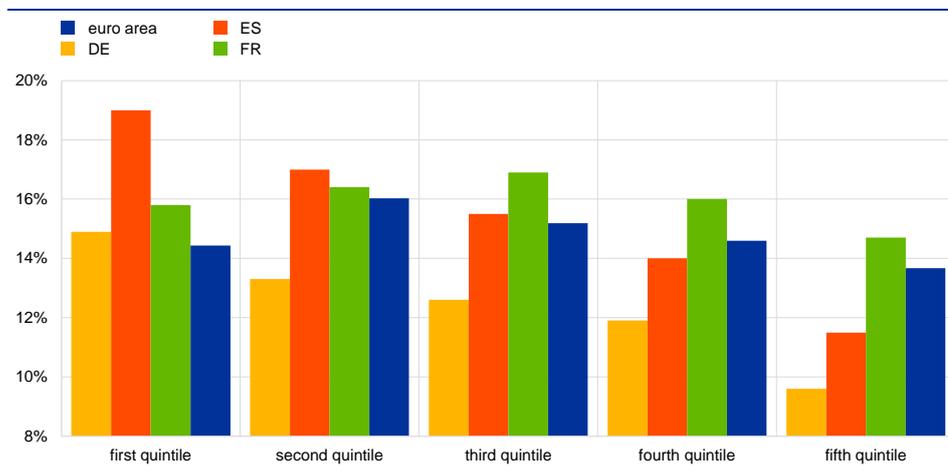
Sources: Eurostat.

The patterns according to other variables may differ across countries. For instance, the share of food increases almost linearly as age increases for the euro area as a whole; it is not the case in DE for instance where it decreases slightly after 60.

2.3 An initial exercise of statistical matching at the macro-level

An initial approach would consist of using averages broken down by some demographic variables that would be common (and comparable) between the different surveys. As an example, it is possible to compute aggregate saving rates broken down by the age of the reference person or by income quintile thanks to the matching of EU-SILC and HBS data for year 2010. This exercise relies on the assumption that the covariates used for the break downs are fully comparable across both surveys. Therefore, when focusing on the age of the reference person, the definition of the reference person is essential in order to ensure full comparability between HBS and EU-SILC data. The retained definition here is the one recommended by the Canberra group (UNECE, 2011). Another source of divergence lies in the point in time taken for defining the age of the reference person: age at the moment of the interview, age at the end of the year, and so on. However, divergences for such a definition should affect only slightly the repartition of households according to the break downs and therefore may be ignored.

Chart 12

Average saving rate broken down by age of the reference person in the euro area
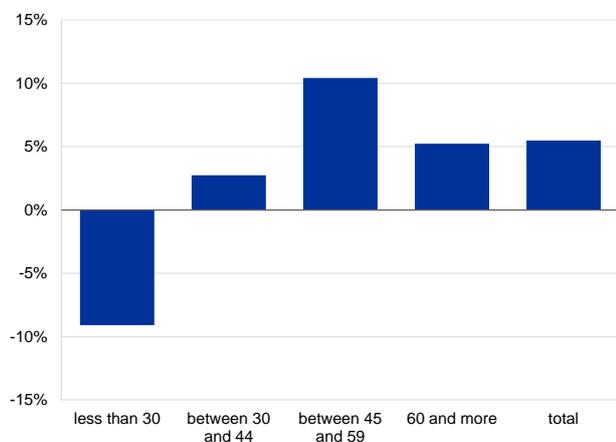
Chart 12 shows the result of this approximate exercise for which average disposable income (as measured in EU-SILC) and average consumption (as measured in HBS) have been broken down into four age groups, which then enables saving rates for each of these groups to be computed. Results are consistent with the data already disseminated; households with a reference person under 30 strongly dissave (the aggregate saving rate for this group of households is -9.1%), then the aggregate saving rate follows a pattern consistent with the life cycle theory: saving rates increase with age, up to 60, then decline in accordance with the drop in income due to retirement. However older people seem on average to keep on saving. This cannot be explained with the basic life cycle theory; dynastic models provide some hints for understanding such a phenomenon.

Another part of the exercise consists of breaking down income and consumption according to income quintile. Indeed, income is very often measured in HBS data, as it may help explain consumption behaviours. However, there is no common definition for income in HBS and it does not necessarily refer to disposable income, making it impossible to compute directly a proper saving rate from HBS data. However, we make a classical "rank" assumption, stating that the household ranking according to income is not sensitive to the definition of income. Hence households in the first quintile of disposable income as measured in EU-SILC are the same as the ones in the first quintile of income as measured in HBS.

Chart 13

Average saving rate broken down by income quintile in the euro area
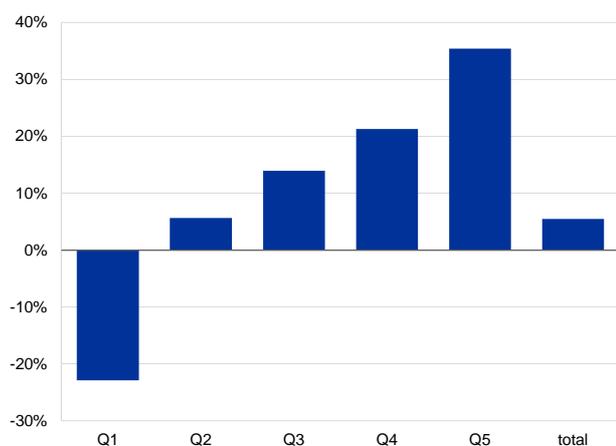
Chart 13 shows the results for such an exercise. Here again, the outcome is consistent with other disseminated exercises: the first quintile of income strongly dissaves (less than -20% in total) whereas the aggregate saving rates increase as income rises.

This exercise makes it possible to exhibit figures easily for some sub-groups of the total population and describe approximately the variety of situations faced by households in terms of budget constraints. However, this work suffers from severe limitations. The first one is related to the comparison with external sources of information, such as National Accounts. Indeed, for the year 2010, the matching between HBS and EU-SILC data shows an aggregate saving rate of 5.5% for the euro area, whereas in the National Accounts this figure is about 13%, almost three times higher than the figure given by surveys. Another limitation is related to the

nature of the data produced: meso-data as such offer very limited possible analyses, as the produced statistics are at most bi-dimensional. This is the kind of drawback that more elaborated statistical matching aims to solve.

3 Methods for estimating consumption at the micro-level

3.1 Overview of the literature on statistical matching

Collecting information on consumption in surveys is a recurrent theme, as it is one key aspect of analysing the economic behaviour of households. Hamermesh (1982) first investigated the question by comparing the measure of consumption as obtained in the Retirement History Survey (RHS) and the one given by the Consumer Expenditure Survey (CE). Those two surveys are aimed at measuring very different dimensions: the RHS follows retired people over time, while the CE is the official survey for measuring American households' expenditure and which is basically the US counterpart of HBS. The author investigated a possible decline in consumption after retirement; taking advantage of the longitudinal data of the RHS, he bases his analysis on the few questions on consumption that exist in this survey for estimating such a decrease, as the CE cannot provide a time span long enough to assess properly the effect of retirement on consumption. Following the same idea, Skinner (1987) links information from the CE about the share of food in total consumption and the value of the main residence with the Panel Study of Income Dynamics (PSID) which aims to describe income in a longitudinal dimension (the US counterpart for the EU-SILC). This exercise allows him to estimate total consumption in the PSID and thereby explain about 78% of the variation of total consumption. This paper constitutes the seminal paper for the method that is intended to be shown in this report.

This framework has been implemented in many surveys, using different approaches that may be more or less complex. One of the most elaborate approaches was described by Blundell et al. (2004) who also estimated consumption in the PSID using information form the CE. They address in a more comprehensive way the issue of potential endogeneity through time of consumption. Indeed, current consumption at least partially results from past decisions of consumption. To do so, the authors use a standard demand function which is the inversion of the equation estimated by Skinner (1987). However, they use a more general model than the one usually used as they also address the potential non-linearity of the demand equation. In a broader perspective, Browning et al. (2003) describe in a comprehensive way all the different methods that may be used for collecting information on consumption in a survey that was originally not intended to focus on such a dimension. In particular, they provide a general framework for the method initiated by Skinner (1987) and they list some generic prescriptions according to the different experiments already implemented in the surveys. These experiments consisted basically of asking questions similar to the ones that are asked in the HFCS. In particular, they analyse results given by questions on broad consumption (that may be called "one-shot questions"), typically worded as follows: "How much do you spend on everything in a typical month?". Such questions have been for instance included in the Canadian Out of Employment (COEP) or the Italian Survey on Household Income and Wealth

(SHIW). They conclude that such questions lead to a strong underreporting phenomenon, which also has been confirmed by Cifaldi and Neri (2013). Nevertheless Browning et al. (2003) also suggest the existence of methods for correcting such biases as long as the underreporting remains consistent over time.

Browning et al. (2003) also comment on the alternative solution of asking questions on a list of precise and limited sub-items; the most satisfying solution consists of having a comprehensive list of sub-items in order to reconstruct easily the total consumption. A recent experiment was conducted with the German Socio-Economic Panel Study (SOEP), with pretty conclusive results (see Marcus, et al. (2013)). This experiment consisted of a module of 16 items that aimed to describe overall consumption. The comparison of the obtained distribution with the one given by the HBS data gives good results; however, adding a complete module in an already long questionnaire increases the response burden substantially. The other option consists of asking questions on specific items that are essential for understanding total consumption: food is one of them. Then a method similar to the one elaborated by Skinner (1987) enables total consumption to be estimated in a general purpose survey. This is a second-best solution; and of course the result of the estimation has to be strongly compared to statistics given by HBS data.

Linking information from two surveys, especially in order to assess links between different variables that do not co-exist in the same survey, is typical of the statistical matching exercise. D'Orazio et al. (2006) propose a general framework for this exercise. Statistical matching is at stake when facing 3 sets of variables X, Y and Z; it is only possible to assess the joint distributions $f_{(X,Y)}$ and $f_{(X,Z)}$ (typically because the surveys measure jointly X,Y and X,Z). They address theoretical and practical issues, not only for estimating the joint distributions $f_{(Y,Z)}$ and $f_{(X,Y,Z)}$, but also for assessing the uncertainty related to such an estimation. The current method set out below is strongly based on the framework they provide. Formal exercises of statistical matching have already been undertaken on survey data in order to link at the same time income, wealth and consumption for European data. For instance, Sutherland et al. (2002) have matched the UK Family Expenditure Survey with the Family Resources Survey in order to run simulations on fiscal policies. Other initiatives may be recalled, such as the one conducted by Eurostat (2013), whose aim was to link information on consumption and on poverty. They have even evaluated the feasibility of matching information between EU-SILC and HFCS data in another report. Cifaldi and Neri (2013) have also linked information for the SHIW with the Italian EU-SILC and HBS. Finally, Baldini et al. (2015) have developed a more sophisticated framework for linking the EU-SILC and HBS data, using a mixture of regressions and conditional hot-deck.

The evaluation of the uncertainty is also essential in order to assess the quality of the estimation and also provide the users with tools that would enable them to have an idea of the validity of the conclusions they may draw from the exercise. Since the estimation is similar with imputation, one solution would consist of computing multiple estimations of the desired variable. D'Orazio et al. (2006) focus on the estimation of uncertainty at the macro-level; they also provide a few references for dealing with uncertainty at the micro-level. For instance, Rubin (1986) describes a

procedure of statistical matching involving multiple imputations; he also addresses in this paper the question of the sample weights that is crucial when performing file concatenation of survey data. As stated by D'Orazio et al. (2006), the exercise of statistical matching often relies on the conditional independence assumption (CIA); this assumption may be strongly questioned, as it assumes the absence of a link between Y and Z conditionally to X. The literature offers different ways for discussing this issue; one most notable answer is the algorithm described by Rässler (2004) that makes it possible to relax the CIA assumption and account for uncertainty over the correlation between Y and Z.

3.2   Description of the main method

3.2.1   Theoretical framework

The estimation of consumption in a general purpose survey may be seen as statistical matching as it involves the estimation of an equation that is based on the information provided by the HBS data. Browning et al. (2003) describe a theoretical framework that relies on very general assumptions about the composition of total consumption. Following this paper, we focus more specifically on the consumption of non-durable goods and services; to justify this choice, Browning et al. (2003) argue that expenditure for non-durable is linked more closely with food consumption than total consumption; in particular, consumption of durable goods relies on very specific determinants and should be established in a different way.

Turning then to non-durable consumption[4], we follow Browning et al. (2003) who considers a list of items $(c_1, \ldots, c_n)$. Each of these n items may be linked with non-durable consumption, assuming a linear Engel curve specification:

$$c_i = \alpha_i + \beta_i c + u_i$$

where c denotes consumption for non-durable and u_i a residual following for instance a normal law. This specification is even more likely to hold with a log transformation. Considering then the share ω_i of item i in non-durable consumption, it is possible to combine the n Engel equations in order to express non-durable expenditure as a linear combination of the different items c_i:

$$c = \left(-\sum_{i=1}^{n} \alpha_i \frac{\omega_i}{\beta_i} \right) + \frac{\omega_1}{\beta_1} c_1 + \cdots + \frac{\omega_n}{\beta_n} c_n - \sum_{i=1}^{n} \frac{\omega_i}{\beta_i} u_i$$

The parameters of this equation may be estimated by OLS on the HBS data and then applied on the HFCS in order to have an estimation of c for each household.

---

[4]   Durable goods are included in items 051, 052, 053, 054, 055, 091, 092 and 093 of COICOP classification. The rest of the items are considered to belong either to non-durable (and semi-durable) goods or services; all are therefore included in non-durable consumption.

3.2.2   Choice of covariates and specification

Once this framework is set, it is essential to determine which items $(c_1, \ldots, c_n)$ have to be included. Following Browning et al. (2003) again, we have to choose the items that are assumed to be accurately measured by simple "recall questions". Items such as "food at home", "food away from home" or "utilities" have turned out to be quite well measured thanks to this kind of question; it is also important that they represent a significant share of non-durable consumption, at least for part of the population. From this point of view, the measurements of the different items from the surveys we are using have to be closely compared, as the comparability of the result may be determinant for the final result. We also focus on the predictive power of these different items over non-durable consumption.

Following the different papers using such a methodology, we also introduce in the equation demographic variables in order to reflect as much as possible the pattern of consumption according to age, level of education, income or household structure. Here again, these variables have to be measured in the most harmonised way between HBS and HFCS data; it is hence essential to compare the different repartitions given by the surveys.

Similarly to Blundell et al. (2004), we adopt a log-specification for the dependent variables and the covariates giving amounts for items of consumption. This log-specification is intended to deal with potential heteroscedasticity of the residuals. Hence the equation that we wish to estimate is the following one:

$$\log(c) = \beta_0 + \beta_1 \log(c_F) + \beta_2 \log(c_O) + \gamma' X$$

where c_F, c_O and X respectively denote the food consumption at home, the food consumption away from home and the different demographic variables. These demographic variables convey some additional information that may be very useful; in particular, it is easy to show that the MPC estimated with such a method will be driven only by food consumption in case there are no demographics in the equation.

3.2.3   Treatment of the uncertainty

Once the equation is estimated, it is therefore possible to compute an estimation of the yearly non-durable consumption for every household belonging to the HFCS. At the household level, the uncertainty should remain quite high in the sense that it is very likely that the consumption for one given household is poorly evaluated. However, the method is rather meant to provide information on consumption for different groups of the population, not only in terms of aggregates, but also in terms of distribution. The uncertainty at the household level is reflected through the residuals that are given by the estimation of the equation in the HBS. Therefore, it is important to take into account this unexplained part, not only in the estimation procedure itself, but also as one key element for the evaluation of the quality of this estimation.

As a first step, we have to consider the procedure for taking into account the residuals in the estimation; we have (at least) three options at our disposal:

- The first one is simply computing the expectancy conditionally to the covariates for each household. As we have assumed that residuals follow a normal law, we can write the expectancy as follows:

$$E(c|c_F, c_O, X) = e^{\beta_0} . c_F{}^{\beta_1} . c_O{}^{\beta_2} . e^{\gamma' X} . e^{\frac{\sigma^2}{2}}$$

- The second one is obtained by drawing residuals from a truncated normal law, with ad-hoc definitions for the lower and the upper bounds (for instance the lower bound may be defined as the sum of the different collected items).

- The last one consists of a conditional hot-deck over the residuals computed on the HBS data. Stratification may help to address potential heteroscedasticity for the estimated equation. This method has been implemented for instance by Cifaldi and Neri (2013).

The uncertainty related to the estimation may be calculated in a first step thanks to a Monte-Carlo algorithm, encompassing the information on uncertainty given by the OLS performed on HBS data. Indeed the parameters of the equation are estimated with an uncertainty which may be assessed thanks to the variance-covariance matrix $\widehat{\Sigma_{\beta,\gamma}}$. The other source of uncertainty is related to the unexplained part of the equation; the dispersion of the residuals and the R² related to the estimation provide an idea of this uncertainty. The uncertainty may be assessed through simulations, thanks to the following algorithm:

- Generate 1,000 coefficients $(\hat{\beta}, \hat{\gamma})$ given that the estimation of (β, γ) follows a normal law of mean (β, γ) and of variance $\Sigma_{\beta,\gamma}$.

- Draw 1,000 times residuals according to the procedure that has been finally applied in order to take into account residuals in the equation.

- Compute 1,000 estimations of consumption which will provide an estimation of the variability of the estimation. The 1,000 estimations may be used for any statistics involving consumption; it may then give an idea of the robustness of the conclusions drawn thanks to this method.

This exercise, as the entire estimation, relies on the assumption that conditionally to the covariates, non-durable consumption and the other variables in the HFCS (e.g. wealth) are independent: this is the Conditional Independence Assumption (CIA) already mentioned before. We then implement the so-called non-iterative Bayesian multiple imputation algorithm (NIBAS) described by Rässler (2004) in order to evaluate how the obtained estimation is sensitive to the CIA. As this is often the case for assessing the range of plausible values in statistical matching, this algorithm relies on the fact that the variance-covariance matrix should be positive definite. For variables X, Y and Z, the CIA means that the correlation between Y and Z conditionally to X is 0. The NIBAS algorithm offers a way of testing all possible correlations between Y and Z and keeping only the ones that provide valid variance-covariance matrices for Y and Z. For a broader description of the algorithm, see

Rässler (2004). However, it relies on strong assumptions about the functional link between consumption and wealth; therefore its results cannot be considered a reliable estimation of the complete range of plausible values for consumption. It only gives an idea of the sensitivity of the estimation, once the price for assumptions different from the CIA is paid.

The quality of the estimation will be assessed thanks to these different algorithms; part of the results that will be presented hereafter are already available in Lydon (forthcoming) and Lamarche (2015).

3.2.4 From non-durable consumption to total consumption

The estimation of non-durable consumption should not hide the ultimate goal of this exercise, which is the estimation of total consumption. Adding consumption of durable goods may not be limited to a methodological issue, but also a conceptual one. On the one hand, depending on the considered good, consumption in the National Accounts may be computed either as a depreciation of the stock of durable goods that is held by households or through the expenses for the purchase of such goods; on the other hand, the HBS data always measure the several purchases of durable goods that households have undertaken during the year. As consumption of durable goods may cover different concepts – although most of the durables that are classified in the HBS are now also seen as expenses in the National Accounts, it may be useful to differentiate between the various uses that may be made from a variable about total consumption.

When it comes to saving rates, the comparison with National Accounts may be essential, as one would be interested in understanding the level of aggregate saving rates in Europe. Moreover, the purchase of durable goods cannot be completely assimilated into consumption in the case of HFCS, as some durable goods may come under the composition of wealth (durables and valuables indeed constitute a component of assets that may be held by households). For these different reasons, it may be interesting to focus on a concept of consumption similar to the one used by the National Accounts. A natural way of computing durable consumption would be hence to break down the flow of aggregate durable consumption between households according to their stock of durables. Considering F, F_i and s_i respectively the aggregate flow of durable consumption, the durable consumption for the household i and the share of durables possessed by the household i, the computation would be written as follows:

$$F_i = F.s_i$$

This method is an approximate way of breaking down aggregate figures at the micro-level. The main advantages are the simplicity of the method and its consistency with concepts and figures coming from the National Accounts. However, it certainly does not reflect a reality at the household level.

If we now focus more on the link between wealth and certain types of consumption (and especially consumption for durables), it would be useful to impute specifically

the purchases for durables that households may have carried out during the year of reference; this necessitates a specific model, as the determinants of such purchases are likely to strongly differ from the ones for non-durable consumption. However, the same demographic variables may enter in the equation, associated with different parameters. Nevertheless, in many cases, one lacks common variables in the HFCS and in the HBS that would be relevant regarding the decision of purchasing durable goods.

3.3 Comparison of the HFCS and the HBS for the participating countries

3.3.1 Definition of the different covariates

As stated by D'Orazio et al. (2006), comparison of the distribution of the covariates is essential in order to assess properly the bias that may affect the estimation in the HFCS. In particular, the variables describing items of consumption have to be relatively close between the HFCS and the HBS.

Three items may be used, depending on their availability in the HFCS: food consumption at home (that is described by variable HI0100), food away from home (variable HI0200) and utilities (variable HNI0100 in HFCS wave 1; variable HI0210 in wave 2). In the HBS data, following the COICOP classification, these different variables are defined as follows:

- Food consumption at home is given by item CP01, including also CP021

- Food consumption away is given by item CP111

- Utilities are given by items CP0443, CP045 and CP0813

Another important item of consumption, especially for those households which do not possess their own home, is rents. Rents here exclude of course imputed rents, even if this specific covariate has been successfully tested for some of the countries (IT). The tenure status of the household is also included in the covariates, as it may strongly affect the consumption behaviour of the household.

Demographic variables are defined according to the reference person. Age, gender, level of education and status of employment are defined on the basis of the reference person. The definition of the reference person in both surveys follows the recommendations expressed by the Canberra group (UNECE, 2011). Hence the reference person is defined as follows:

(a) One of the partners in a registered or de facto marriage, with dependent children,

(b) One of the partners in a registered or de facto marriage, without dependent child

<pre>   (c)   Alone parent with dependent children

        •   The person with the highest income,

        •   The eldest person.
</pre>

The size of the household and the number of children are also encompassed in the covariates, as they do not pose any conceptual issues. Finally, income is also included, as it may strongly influence behaviours in terms of consumption. However, there are different comparability issues regarding income; in most cases, incomes are not directly comparable between the HFCS and the HBS. In most cases, the concepts of income differ, HFCS focusing on gross income while HBS data encompass information on net income. Also, the modes of collection are very often different. All these different issues make the use of such a covariate difficult. One way of dealing with these different issues is rather to use income quintiles, thereby making the assumption that the hierarchy across the various ways of measuring income remains the same.

3.3.2   Comparing distribution

As stated in Browning et al. (2003), the comparability between the different sources that are meant to be matched is essential. Even if the first wave of the HFCS is considered to be 2010, the reference periods may vary across countries, starting from 2008 for ES to 2011 for DE, LU, MT and AT. Therefore, it may be good to choose the HBS wave that better fits the reference period for the HFCS. As in some cases, the HBS occurs every year, it is possible to estimate the equation using very appropriate data; in other cases the HBS only occurs every 5 years. The reference period in the European Union is set to 2010; however there may be some differences that in the end may have an effect on the results. At any rate, the comparability of both sources has to be assessed, comparing the distributions of the covariates according to HBS and HFCS data.

First of all, as food consumption explains about two thirds of the variation of total consumption (**Table 2**), it constitutes a key variable for which the comparability is essential. Also the risk of discrepancy for this specific variable is higher than for the other covariates: indeed the way this variable is collected differs widely between the two surveys. In the case of HBS, information on expenditure for food consumption (away or/and at home) is gathered using diaries that households are asked to fill in during the fieldwork. In the case of the HFCS, food consumption is addressed through only two questions, distinguishing consumption at home from consumption away from home. These two different ways of collecting information are likely to give on average (and also in terms of variability) very different results. This is the reason why one should carefully check the distribution before using it as a covariate.

Chart 14

Estimated distributions for food consumption for BE
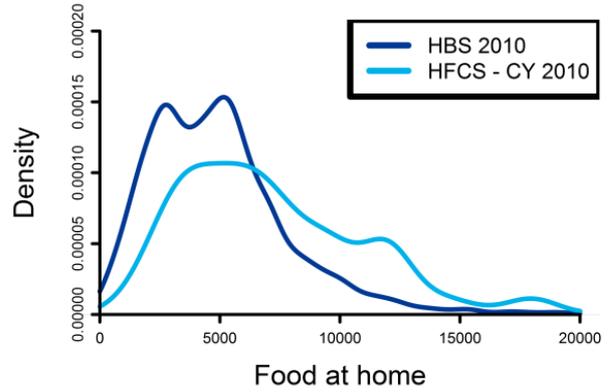
(kernel density estimates; optimal Gaussian kernel)
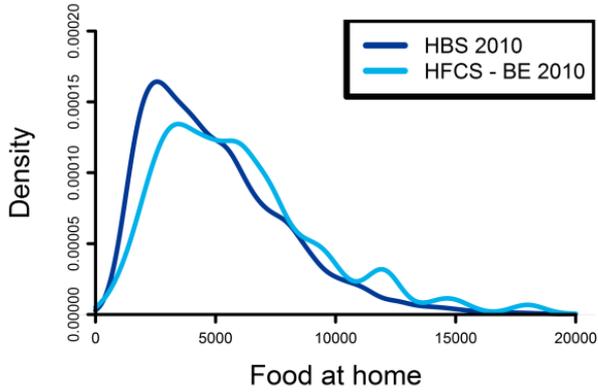
Sources: HFCS and HBS, author's computations.

Chart 15

Estimated distributions for food consumption for CY

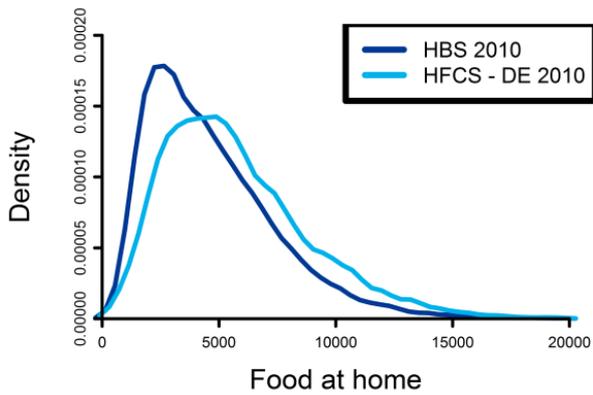(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 16

Estimated distributions for food consumption for DE
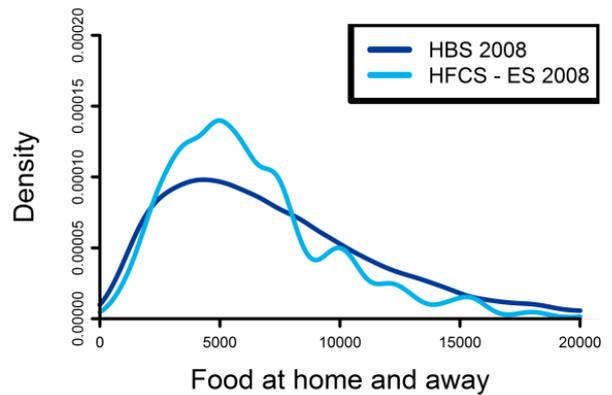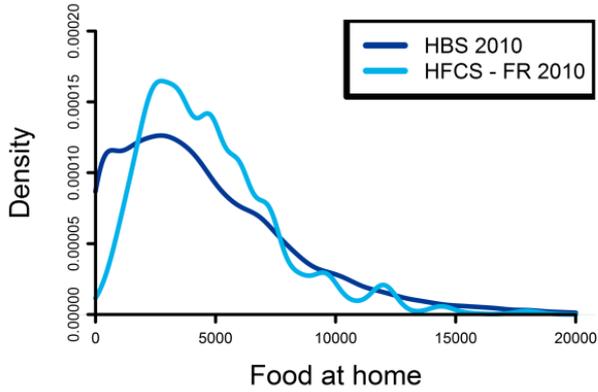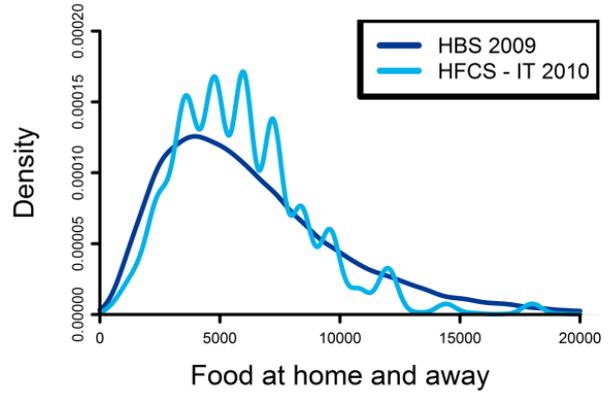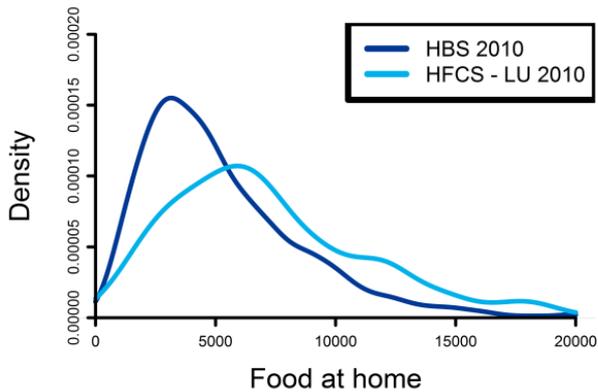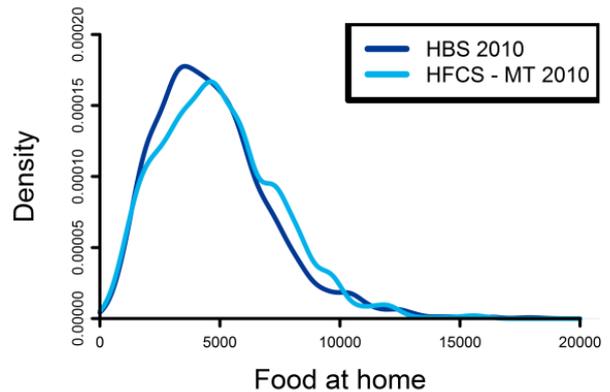
(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, computation by Bundesbank.

Chart 17

Estimated distributions for food consumption for ES

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 18

Estimated distributions for food consumption for FR

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 19

Estimated distributions for food consumption for IT
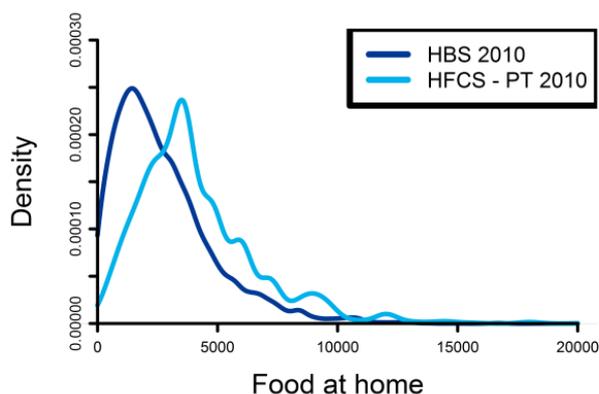
(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 20

Estimated distributions for food consumption for LU

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 21

Estimated distributions for food consumption for MT

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 22

Estimated distributions for food consumption for PT
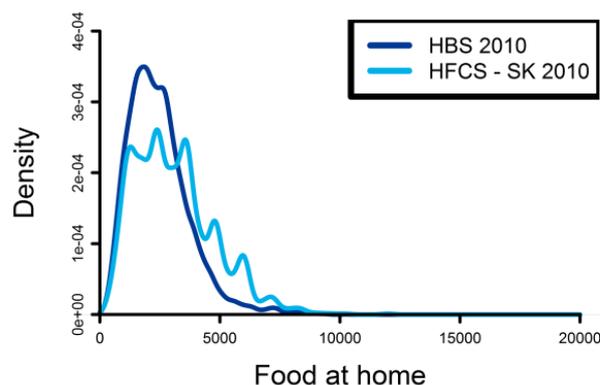
(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, computations by Banco de Portugal.

Chart 23

Estimated distributions for food consumption for SK

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

As shown from **Chart 14** to **Chart 23**, there are strong differences between the different measures of food consumption. The global shape of distribution is somewhat preserved in general; however the percentiles differ sometimes strongly. On average also, the results are not the same across the different surveys. Finally, it is worth noting that there is a strong rounding effect as the amounts given by households when asked one single question is most of the time rounded, whereas they are not when the information is collected through diaries.

Another potential factor for explaining non-durable consumption could be the utilities. Regarding the variable measured in the HFCS it is a non-core variable for the first wave and it is available only for FR (one third of the sample). For the second wave, it is a core variable and has been used for instance for estimating consumption for IE (see Lydon, forthcoming). In IE, the HFCS overestimates on average expenditure for utilities by 39%; in FR, the mean is overestimated by 53%. According to French results, this overestimation is overall the consequence of the overestimation at the top of the distribution; the median in the HFCS is 14% higher than the one given by HBS.

Regarding rents, the comparison between HBS and HFCS data gives pretty good results. First of all, in most cases, the repartition for the tenure status is similar, as shown in **Table 9** in annex. With the exception of FR, the distribution of the rents conditionally to being tenant is rather close between HBS and HFCS. In the case of IT, the imputed rents have also been included, as they show consistent distributions across both surveys.

Regarding demographic variables, the common variables show pretty comparable repartitions (see **Table 9**), as they have often been collected in the same way in both surveys.

Another good indication of the similarity of the two surveys would be the resemblance in terms of variability, regardless of differences in terms of scale. In

Chart 24, we compare the coefficients of variations for food consumption as measured in the HFCS and in the HBS broken down by income quintile.

Chart 24
Ratios between coefficients of variation of food consumption according to the income quintile
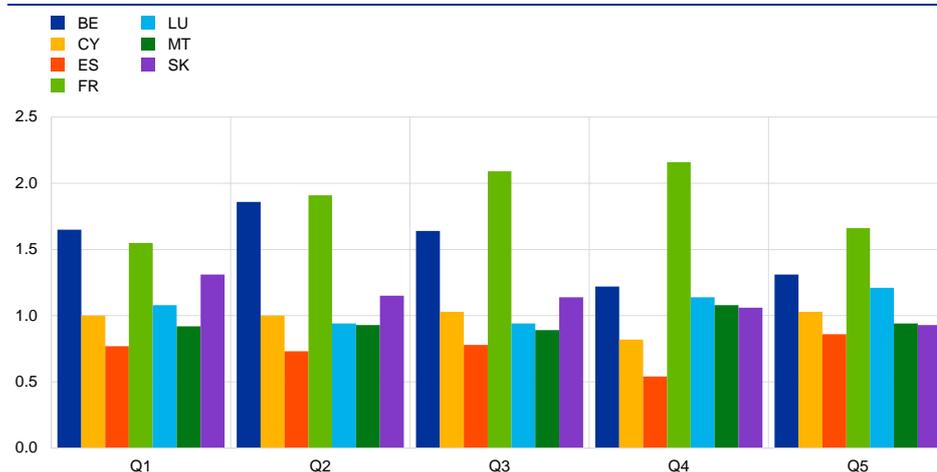
Sources: HFCS and HBS, author's computations. A ratio less than 1 indicates that the coefficient of variation in HBS is higher than the one in the HFCS.

First of all, it turns out that the difference in terms of variability is not the same depending on the country. For instance, for FR HFCS data show higher CVs than in HBS (74% higher); in ES or in IT, the CVs in the HFCS are lower. More importantly, the difference of variability may vary across age, level of education, size of the household or income (see **Chart 24** and **Table 10**). This result is of concern, as it reveals that the final result may be distorted in some way within the different classes defined by the demographic covariates.

4 Results for the different countries

4.1 Explanatory power of the covariates

Apart from the considerations on the comparability of the variables, the explanatory power of the different variables should be assessed, as it reveals the likelihood of a given item of consumption reflecting properly the evolution of total consumption.

Table 2

Explanatory power of the models

(OLS - Adjusted R²)

Models	BE	CY	DE	ES	FI	FR	IT	LU	MT	SK
Only food consumption	0.60	0.71	n.a.	0.66	0.57	0.61	0.51	0.61	0.65	0.70
Only demographics	0.44	0.57	n.a.	0.48	0.56	0.44	0.36	0.36	0.37	0.54
Food consumption, demographics	0.67	0.80	0.67	0.73	0.71	0.71	0.60	0.69	0.69	0.78
Food consumption, demographics, income dummies	0.70	0.82	0.71	0.75	0.75	0.77	n.a.	0.72	0.71	0.81
Food consumption, demographics, income dummies and interactions	0.70	0.83	n.a.	0.76	0.77	0.78	n.a.	0.72	0.72	0.81

Sources: HBS. Computation by Bundesbank for DE, author's computations for other countries.

As shown in **Table 2**, the inclusion of food consumption in the equation improves significantly the explanatory power of the covariates for describing non-durable consumption. There is then a trade-off between the improvement of the explanatory power of the equation and the noise and bias that the inclusion of this variable may add, as the mode of collection differs strongly between the two surveys.

The other crucial point is the inclusion or not of the income in the equation. One initial choice that has been made for this work is to use the quintile of income in order to eliminate potential strong discrepancies between the two surveys for measuring income. Indeed, depending on the survey, the mode of collection but also the definition of income may strongly vary: gross or net income, collected through one single question, a set of questions describing the different components, or retrieved thanks to administrative data, the variable may suffer from various biases and distortions. Therefore, including income is possible thanks to the definition of a hierarchy with respect to income, thereby making the assumption that the global ranking is not sensitive to either the definition of income or its mode of collection – the 'rank assumption' mentioned in section 2.3. However, including income does not turn out to be straightforward: on the one hand, Browning et al. (2003) point out that the inclusion of such a variable in the model may introduce potential spurious relationships between income and consumption. On the other hand, one would argue that broad conditioning would justify the inclusion of any variable that may have a relevant effect on consumption in the equation. A pragmatic approach consists of comparing the results with and without income; as shown in **Table 2**, the inclusion of the variable improves slightly the explanatory power of the equation. However, before arguing for its inclusion or its exclusion, it is also necessary to look at the other criteria for assessing the quality of the imputation.

4.2 Comparing the distribution of non-durable consumption between observed HBS and estimated HFCS data

Once the estimation of the equation made on the HBS data has been made, the second natural step consists of applying the estimated coefficients to the HFCS data in order to estimate non-durable consumption. As explained in section 3.2.3, one key point of this work is the inclusion of the residuals. Different options have been tested, with no clear-cut preference for one or the other.

Different sets of covariates have been tested, especially regarding the inclusion or not of income variables. The outcome of part of the comparative analysis is shown from **Chart 25** to **Chart 32**, for which the density of consumption estimated with or without income is shown. It is also possible to compute unidimensional statistics which would then provide an unambiguous hierarchy between the different options, such as the statistics for the Kolmogorov-Smirnov test. However, including income turns out, from a graphical point of view, not to have a strong impact on the final estimation. In the end, the argument on broad conditioning should prevail, as conditioning with income is more likely to preserve the already observed correlation between income and consumption.

Chart 25

Estimated distributions for non-durable consumption in BE - including and not including income as a covariate
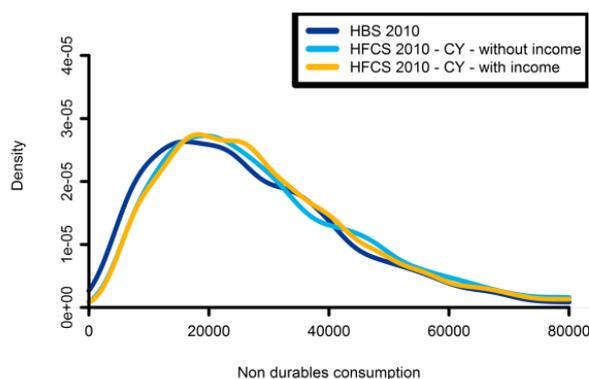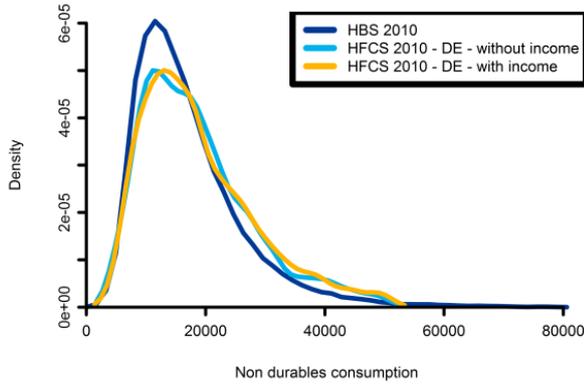
(kernel density estimates; optimal Gaussian kernel)
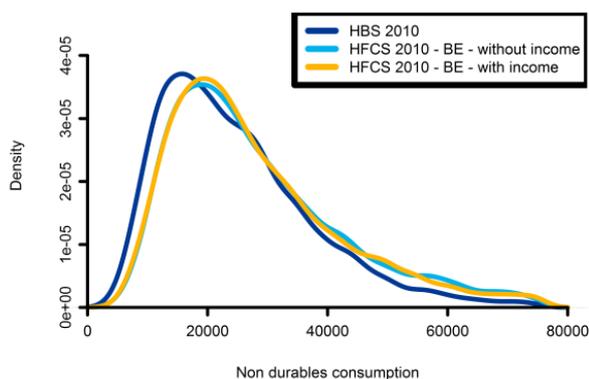
Sources: HFCS and HBS, author's computations.

Chart 26

Estimated distributions for non-durable consumption in CY - including and not including income as a covariate

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 27

Estimated distributions for non-durable consumption in DE - including and not including income as a covariate

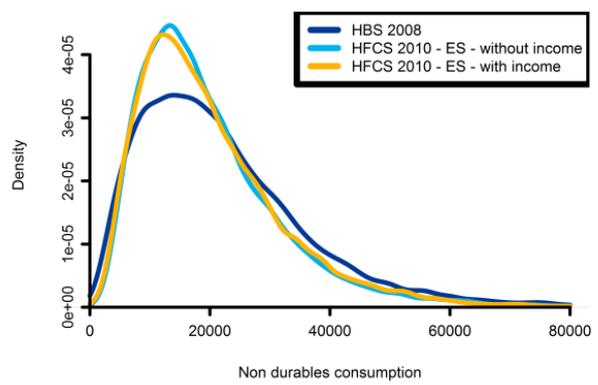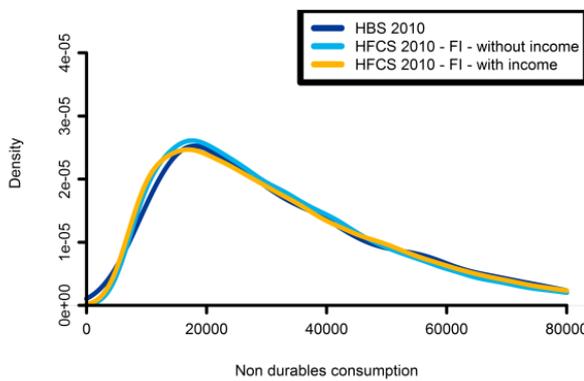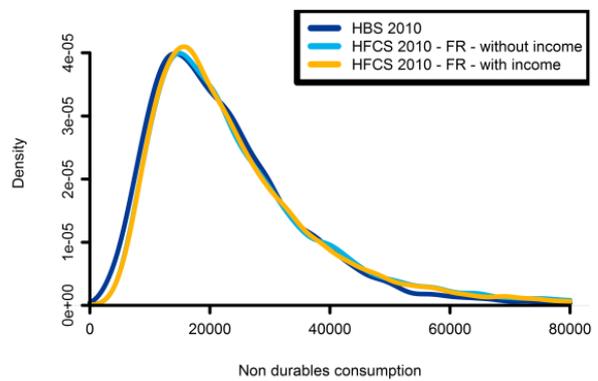(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, computation by Bundesbank.

Chart 28

Estimated distributions for non-durable consumption in ES - including and not including income as a covariate

(Kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 29

Estimated distributions for non-durable consumption in FI - including and not including income as a covariate

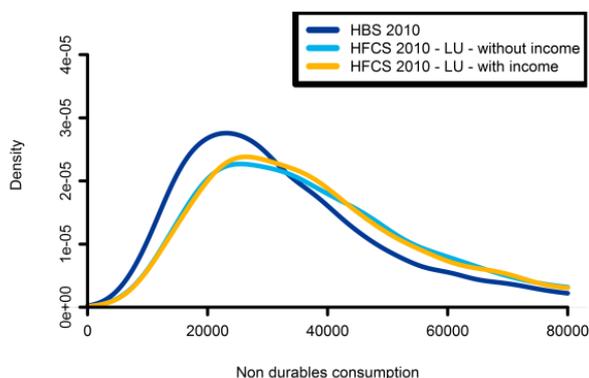(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 30

Estimated distributions for non-durable consumption in FR - including and not including income as a covariate

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 31

Estimated distributions for non-durable consumption in LU - including and not including income as a covariate

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 32

Estimated distributions for non-durable consumption in MT - including and not including income as a covariate

(kernel density estimates; optimal Gaussian kernel)
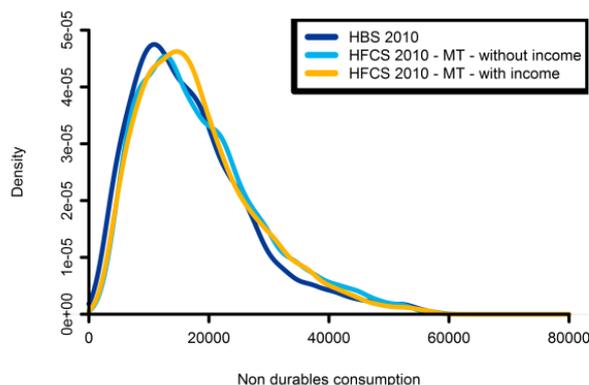
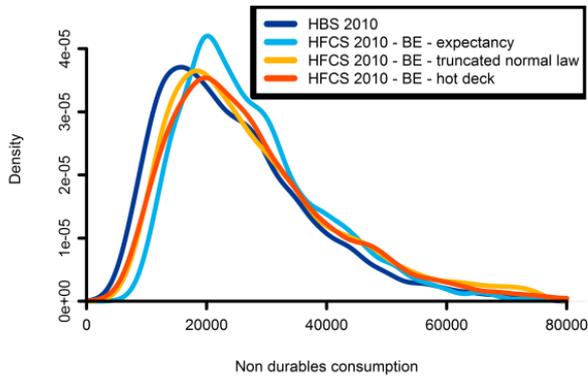Sources: HFCS and HBS, author's computations.

Different specifications including interaction between income variables and other covariates have also been tested. In particular, interactions between food consumption and income have been tested in order to take into account potential Engel law's effects. Indeed, even a log specification with polynomial terms may not be sufficient to address potential non-linearity in the upper part of the income distribution. **Table 11** shows the final specification for taking into account interactions between income and food consumption.

Regarding food consumption, some countries have collected the information in one single variable, pooling altogether consumption at home and away from home. Some others have collected the two variables separately. It is then possible to test across this set of countries models that distinguish between the two variables or not: a Chow test makes it possible to conclude that it is preferable to separate food consumption at home and away from home.

Chart 33

Estimated distributions for non-durable consumption in BE - specification of the residuals
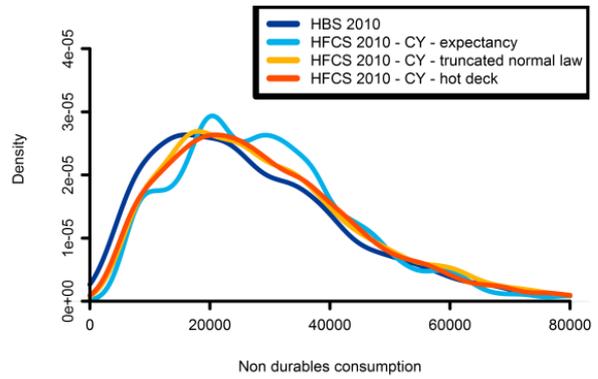
(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 34

Estimated distributions for non-durable consumption in CY - specification of the residuals
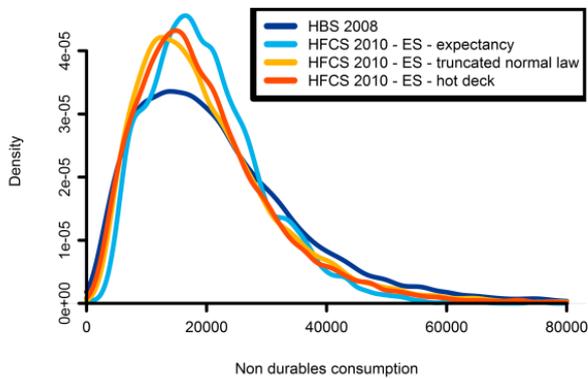
(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 35

Estimated distributions for non-durable consumption in ES - specification of the residuals
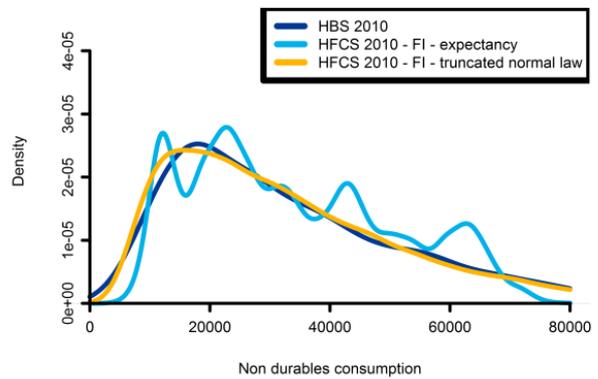
(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 36

Estimated distributions for non-durable consumption in FI - specification of the residuals

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 37

Estimated distributions for non-durable consumption in FR - specification of the residuals

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 38

Estimated distributions for non-durable consumption in IT - specification of the residuals

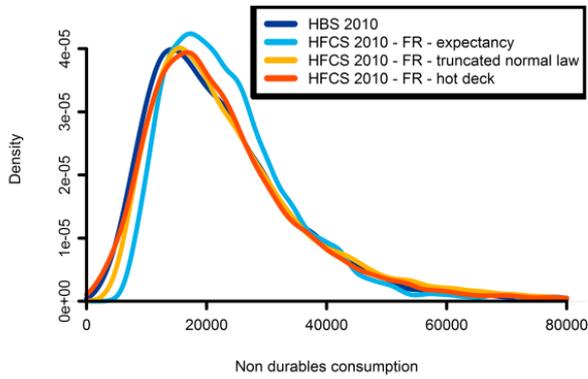(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 39

Estimated distributions for non-durable consumption in LU - specification of the residuals

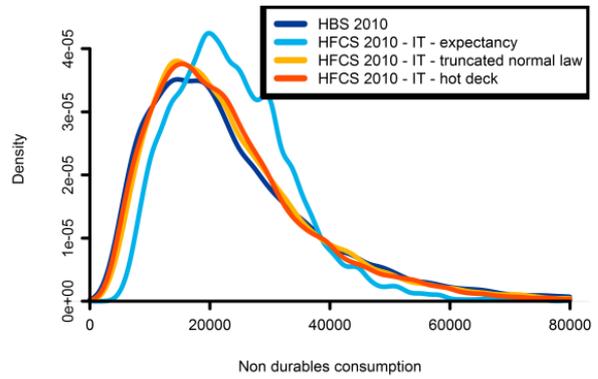(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 40

Estimated distributions for non-durable consumption in IT - specification of the residuals

(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Chart 41

Estimated distributions for non-durable consumption in SK - specification of the residuals

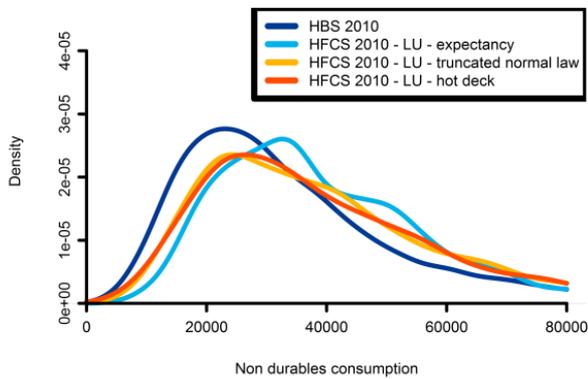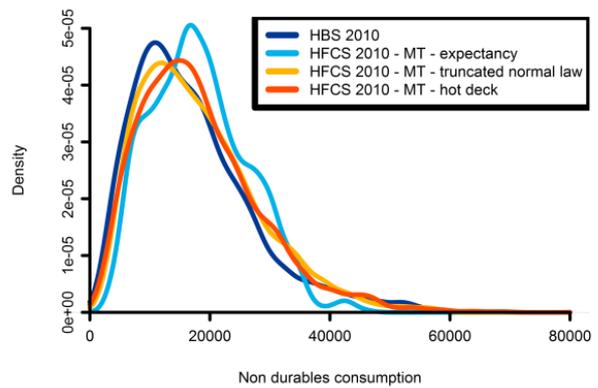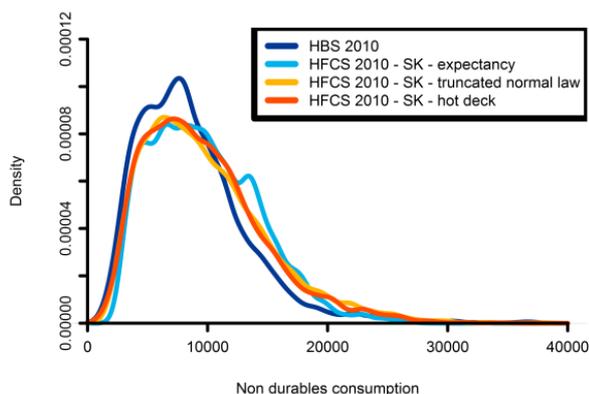(kernel density estimates; optimal Gaussian kernel)

Sources: HFCS and HBS, author's computations.

Finally the question of the residuals to be added to the consumption equation has to be tackled, as it may have an impact on the final estimation. Different methods for taking into account the residuals have been implemented; the most simple consists of computing the expectancy for each household conditionally to his characteristics. This method requires very little information apart from the covariates; however, it suffers from two strong weaknesses. The first one is related to the fact that the set of possible values to be attributed to the household may then be very limited, leading to accumulation points in the distribution: this is for instance particularly striking for Finnish data (see **Chart 36**). The second one is linked with the normality assumption, which may be questionable particularly in the case of heteroscedasticity.

There is then a choice to be made between the conditional hot-deck and the truncated normal law. The main drawback of drawing residuals from a truncated normal law is that here again we make a strong assumption, whereas hot-deck only requires residuals from the regression performed on HBS data. This last option has the advantage of being assumption-free; however, one needs to disseminate a derived form of the HBS micro-data in order to enable the user to replicate the exercise. The final results obtained with a truncated normal law are given in **Table 3**; they can also be computed by the reader in SAS or Stata thanks to the scripts that are in annex of this paper.

Table 3

Ratios between percentiles as measured in HBS and the ones estimated in HFCS

(chart information)

Indicators	BE	CY	DE	ES	FI	FR	IT	LU	MT	PT	SK
P1	1.25	1.43	1.01	1.46	0.98	1.42	1.32	1.24	1.33	1.75	1.19
P5	1.19	1.36	1.06	1.18	0.92	1.19	1.30	1.16	1.28	1.38	1.10
P10	1.15	1.26	1.11	1.09	0.94	1.11	1.30	1.17	1.18	1.27	1.11
P25	1.13	1.14	1.12	0.97	0.96	1.07	1.20	1.18	1.08	1.11	1.10
P50	1.09	1.05	1.13	0.90	0.99	1.05	1.16	1.20	1.07	1.03	1.13
P75	1.10	1.00	1.12	0.87	0.97	1.06	1.06	1.17	1.04	1.01	1.15
P90	1.14	0.98	1.16	0.87	0.96	1.14	1.06	1.16	1.01	1.03	1.16
P95	1.16	0.99	1.18	0.86	0.97	1.26	1.03	1.22	0.94	1.09	1.15
P99	1.22	1.00	1.35	0.82	0.94	1.92	0.99	1.43	0.87	1.40	1.06

Sources: HBS and HFCS data. Computations by Bundesbank for DE, computations by Banco de Portugal for PT, author's computations for other countries.

4.3 Comparison of break downs for non-durable consumption between HBS and HFCS data

In order to assess the quality of the final estimation of consumption in the HFCS, it is possible to compare the different breakdowns obtained with this estimation with the ones given by the HBS. As the breakdowns are made with variables that are included in the estimation equation, we expect the results to be close to what we observe in HBS. However, significant gaps between HFCS and HBS data would indicate that the estimation failed in reproducing expected patterns for consumption. We investigate break downs according to the age of the reference person, his or her level of education, the structure of the household and the income quintile. We specifically focus on three different indicators: the median, the mean and the share of food consumption.

Chart 42
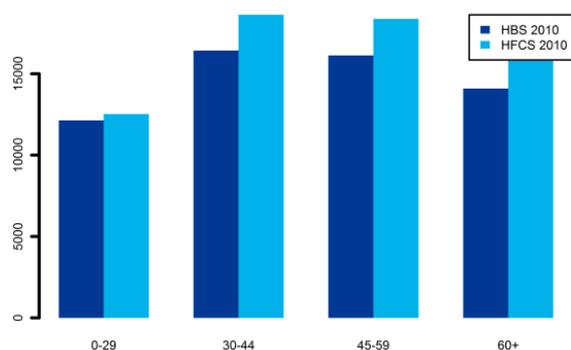
Comparison of medians by age of RP in BE
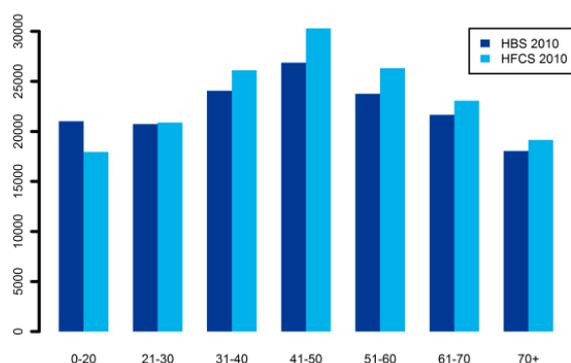
Sources: HFCS and HBS, author's computations.

Chart 43
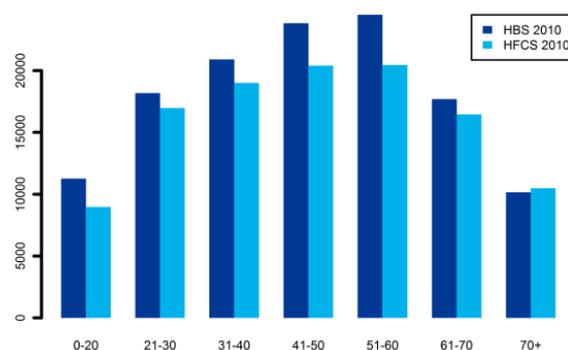
Comparison of medians by age of RP in CY
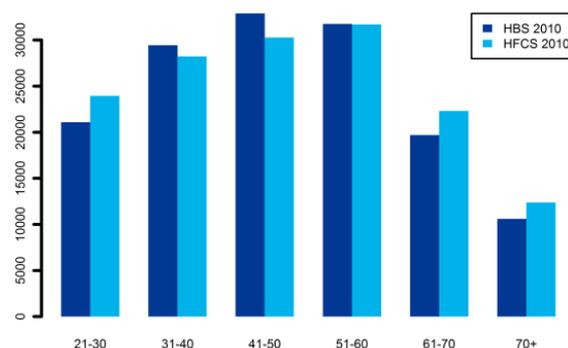
Sources: HFCS and HBS, author's computations.

Chart 44

Comparison of medians by age of RP in DE

Sources: HFCS and HBS, computations by Bundesbank.

Chart 45

Comparison of medians by age of RP in ES

Sources: HFCS and HBS, author's computations.

Chart 46

Comparison of medians by age of RP in FI

Sources: HFCS and HBS, author's computations.

Chart 47

Comparison of medians by age of RP in FR

Sources: HFCS and HBS, author's computations.

Chart 48

Comparison of medians by age of RP in IT

Sources: HFCS and HBS, author's computations.
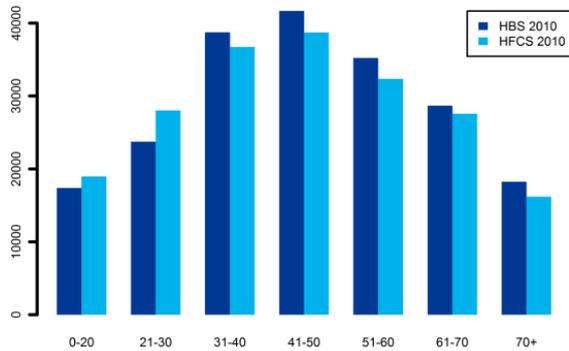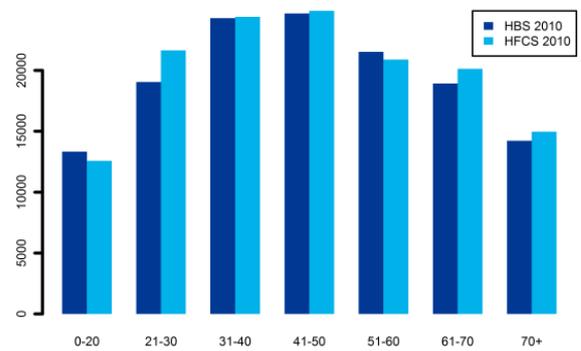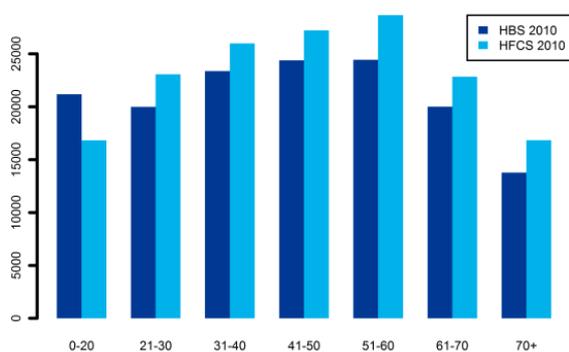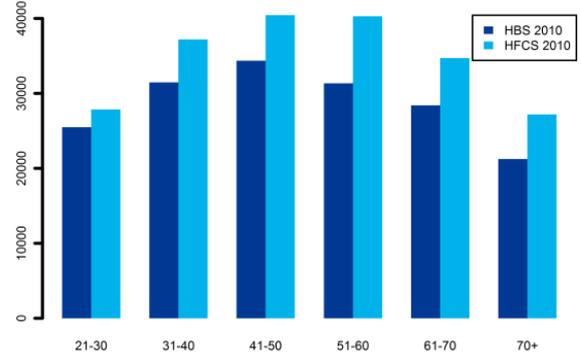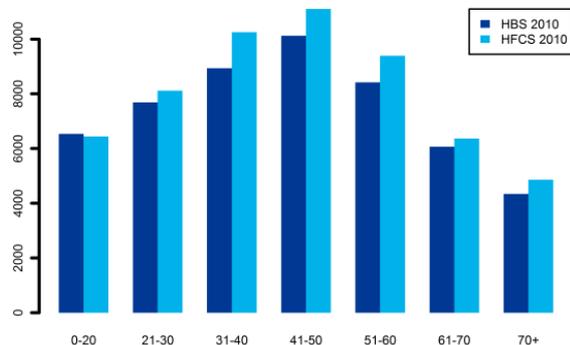
Chart 49

Comparison of medians by age of RP in LU

Sources: HFCS and HBS, author's computations.

Chart 50

Comparison of medians by age of RP in SK

Sources: HFCS and HBS, author's computations.

As shown from **Chart 42** to **Chart 50**, the medians computed for each class of age in the HFCS are relatively close to the ones observed in the HBS. At least the pattern according to the age for consumption turns out to be the same in the two surveys.

One other crucial point is the replication of the pattern for the share of food as part of consumption according to age. The share of food may decrease after 30 and then increase when the household's reference person is more than 44: this is particularly the case for IT. However, section 2.2 has proved that for euro area as a whole, the share of food rather increases as age rises. Finally, for several countries, the models turn out not to reproduce in a satisfactory way the expected pattern (**Table 4**).

Table 4

Comparison of food as a share of consumption, between HBS and HFCS data according to the age of the reference person

Age	BE	CY	DE	ES	FR	IT	LU	SK
0-29	1.9	5.0	4.4	-3.0	0.7	-2.6	3.4	1.3
30-44	2.2	5.6	4.8	-4.4	2.2	0.2	2.3	4.8
45-59	2.5	10.3	4.8	-1.8	0.8	-0.1	2.2	3.8
60+	3.5	10.3	5.5	0.9	1.6	-1.0	5.6	0.7
Total	2.5	8.5	4.9	-1.9	3.4	-0.6	3.2	3.1

Sources: HBS and HFCS data. Computation by Bundesbank for DE, author's computations for other countries.

Also in annex (from **Chart 75** to **Chart 84**), medians broken down by the household structure are quite comparable between HBS and HFCS, although sometimes the estimated consumption in the HFCS fails to reproduce the variance between the different categories. Particularly for IT, the model performs badly with respect to the household structure.

Table 5

Comparison of food as a share of consumption, between HBS and HFCS data according to the household structure

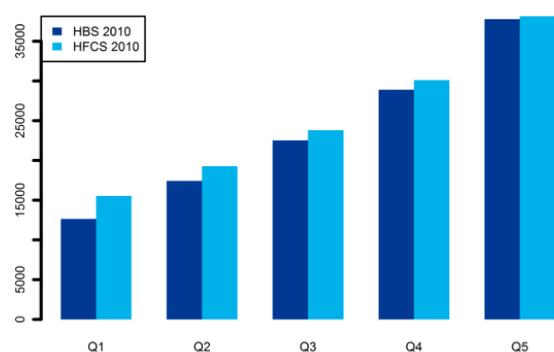| Type	BE	CY	DE	ES	FR	IT	LU	MT	SK
1 person	4.3	7.1	6.5	1.5	3.5	-0.9	7.7	-5.9	-0.6
1 person with children	3.4	5.9	5.7	-1.2	0.2	-2.2	4.6	-4.5	3.3
2 persons	2.9	6.4	5.0	-0.9	-0.3	-0.6	7.4	-5.7	0.1
2 persons with children	2.1	8.2	1.0	-3.0	1.5	-0.3	5.8	-3.9	5.5
3+ persons	-4.5	12.8	4.2	-2.0	3.0	-0.9	6.7	-4.0	3.9
3+ persons with children	3.8	10.5	6.6	-4.0	0.8	-3.0	4.5	-3.9	5.9
Total	2.5	8.5	4.9	-1.9	3.4	-0.6	3.2	0.7	3.1

Sources: HBS and HFCS data. Computation by Bundesbank for DE, author's computations for other countries.

Conversely, the share of food is rather stable in IT according to the household structure. Indeed, it varies between 26% and 30%, while the estimation in the HFCS shows greater variations. The pattern is satisfyingly reproduced for ES; it is not the case for FR (**Table 5**).

Finally, the results are also compared according to the level of education of the reference person and income. Income is sometimes not available (as is the case for IT) and therefore it is not always possible to perform this kind of comparison.
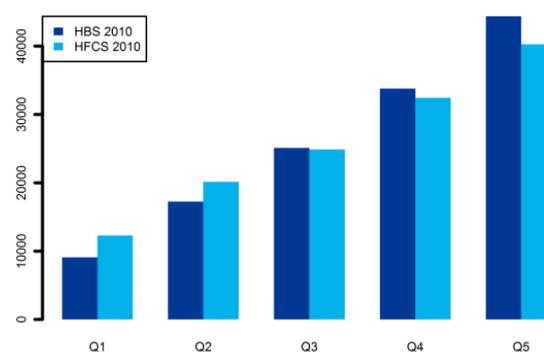
Chart 51

Comparison of medians by age of RP in BE

Sources: HFCS and HBS, author's computations.

Chart 52

Comparison of medians by age of RP in CY

Sources: HFCS and HBS, author's computations.

Chart 53

Comparison of medians by age of RP in DE

Legend: HBS 2010, HFCS 2010
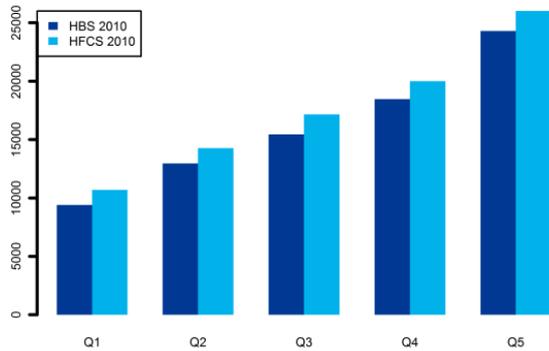
Sources: HFCS and HBS, computations by Bundesbank.

Chart 54

Comparison of medians by age of RP in ES

Legend: HBS 2008, HFCS 2010

Sources: HFCS and HBS, author's computations.

Chart 55

Comparison of medians by age of RP in FI

Legend: HBS 2010, HFCS 2010

Sources: HFCS and HBS, author's computations.

Chart 56

Comparison of medians by age of RP in FR

Legend: HBS 2010, HFCS 2010

Sources: HFCS and HBS, author's computations.

Chart 57

Comparison of medians by age of RP in LU

Chart 58

Comparison of medians by age of RP in MT

Sources: HFCS and HBS, author's computations.

Sources: HFCS and HBS, author's computations.

Chart 59

Comparison of medians by age of RP in SK

Sources: HFCS and HBS, author's computations.

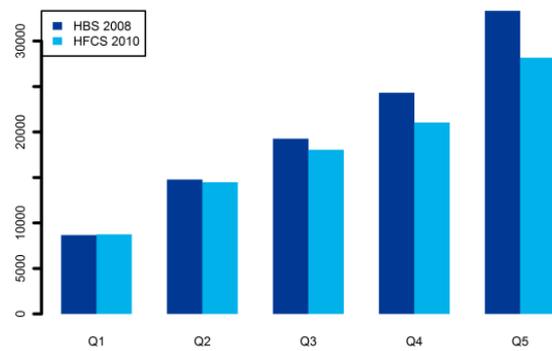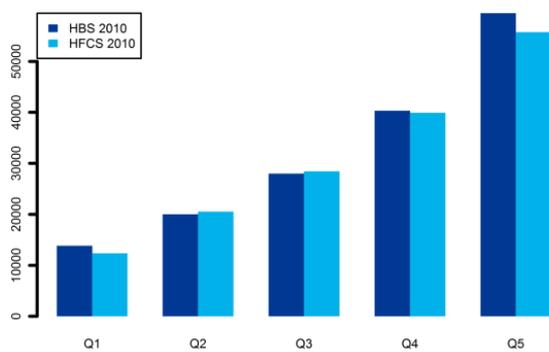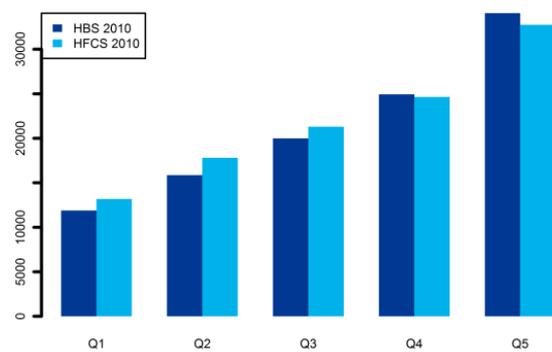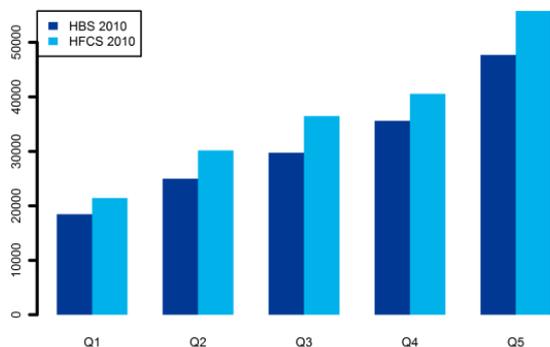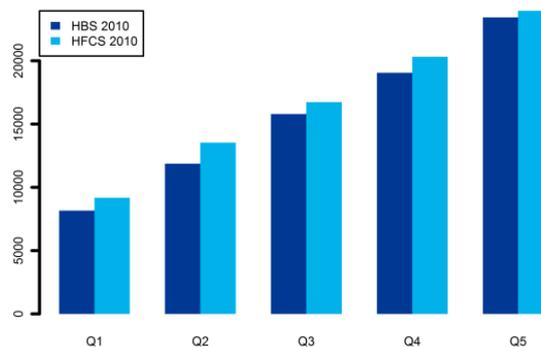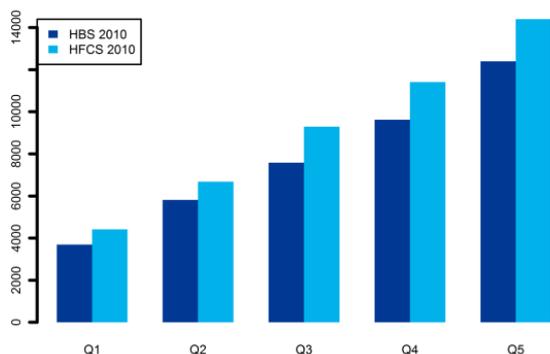The compared medians for ES and FR broken down by income quintiles are shown from **Chart 51** to **Chart 59**. As expected, the median consumption is increasing with the quintile; however the gap sometimes increases with the level of income, especially regarding ES and CY; the inverse holds for LU and SK. Here again, the estimation does not reflect completely the variance of consumption between the different quintiles of income. It is also useful to recall here that, as stated in section 3.2.2, the inclusion of income quintiles in the model relies on an assumption of ranking invariance, although there are great differences in terms of income levels between HBS and HFCS.

Regarding the share of food, contrary to the Engel's law, French HBS data show a stable share of food as part of consumption. Such a result has already been underlined in section 2.2. The estimation reproduces this pattern, as the share of food in the HFCS remains almost stable along the income distribution, around 21% in the first and in the last quintiles. For ES, the results are more clear-cut, especially because the HBS data respect Engel's law: HBS data show that the share of food decreases from 37% in the first quintile to 32% in the last quintile. As a consequence, the estimation in the HFCS follows a similar path, as the share of food drops from 36% in the first quintile to 30% in the last quintile.

Finally, regarding the level of education, the results are similar to the ones obtained for income. This is quite expected, as the level of education may be regarded as a proxy for income. The most interesting results concern IT, for which we do not have any result on income. The median doubles between the low-educated people to people with a tertiary level (from EUR 14,000 to EUR 28,000 according to HBS data). This is not the case in the HFCS data where the median remains stable, around EUR 20,000. As for the share of food, it is expected to decrease as the level

of education goes up; this is observed in the estimated HFCS data, where the share decreases from 26% for the less educated people to 19% for the most educated ones.

4.4 Assessment of the uncertainty

4.4.1 First approach: a basic Monte-Carlo exercise

As described in section 3.2.3, it is possible to assess the uncertainty related to the estimation of consumption in the HFCS, taking as an initial approach the uncertainty due to the estimation of the coefficients and the unexplained part of the equation. Simulating both the parameters of the equation and the residuals makes it possible to take into account the different sources of uncertainty, assuming that the Conditional Independence Assumption (CIA) still holds. Such a simulation is then implemented in order to compute a significant number of estimations for consumption (in this case, we compute 1,000 estimations). It is then possible to use the classical inference tools for estimating the confidence intervals related to our estimation.

As shown from **Chart 60** to **Chart 68**, the confidence interval for the estimation of the distribution of non-durable consumption is rather small.

Chart 60

Confidence intervals for the percentiles of non-durable consumption in BE
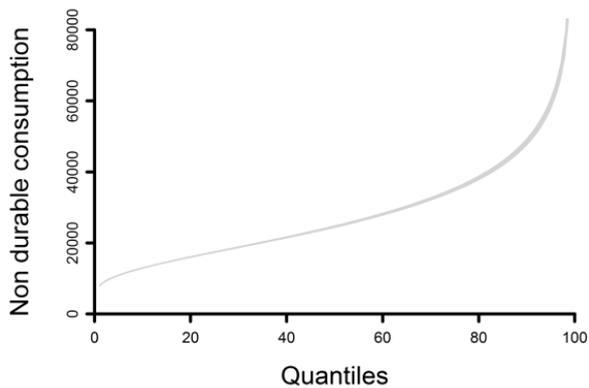
Sources: HFCS and HBS, author's computations. Confidence intervals at 95%.

Chart 61

Confidence intervals for the percentiles of non-durable consumption in CY

Sources: HFCS and HBS, author's computations. Confidence intervals at 95%.

Chart 62

Confidence intervals for the percentiles of non-durable consumption in ES

Sources: HFCS and HBS, author's computations. Confidence intervals at 95%.

Chart 63

Confidence intervals for the percentiles of non-durable consumption in FI

Sources: HFCS and HBS, author's computations. Confidence intervals at 95%.

Chart 64

Confidence intervals for the percentiles of non-durable consumption in FR

Sources: HFCS and HBS, author's computations. Confidence intervals at 95%.

Chart 65
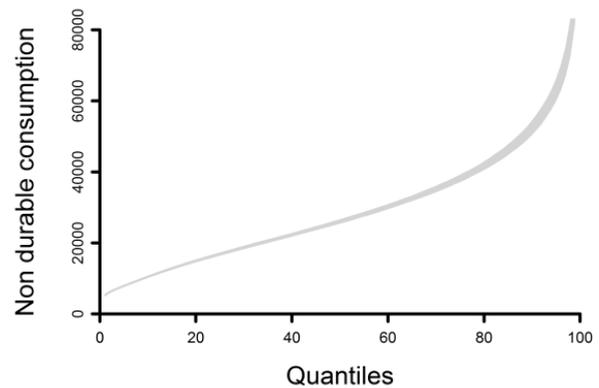
Confidence intervals for the percentiles of non-durable consumption in IT
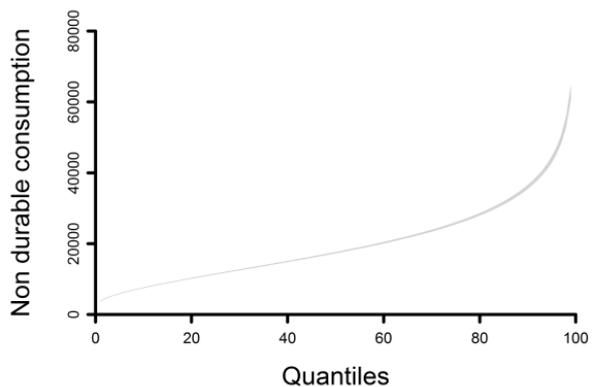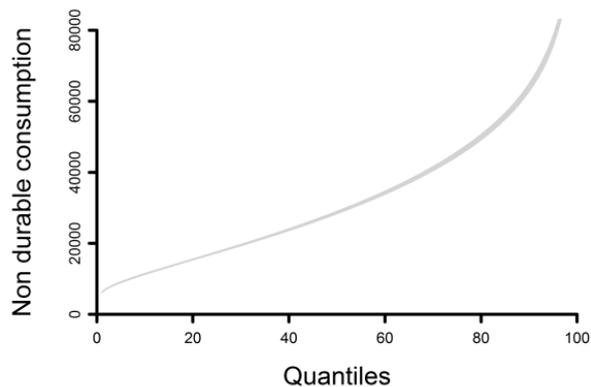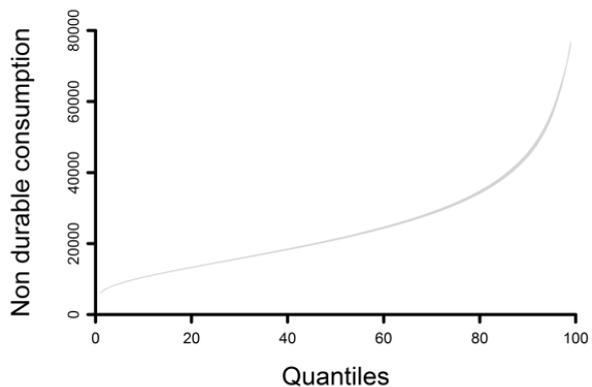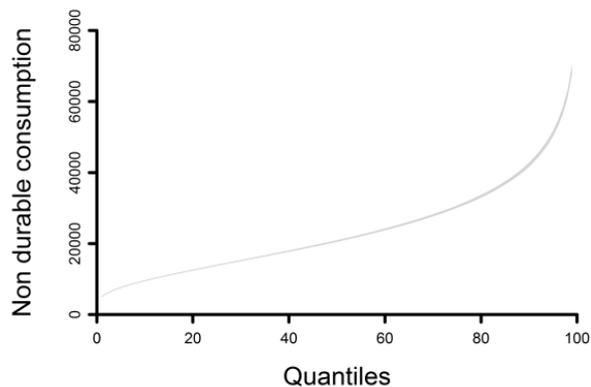

Sources: HFCS and HBS, author's computations. Confidence intervals at 95%.

Chart 66

Confidence intervals for the percentiles of non-durable consumption in LU

Chart 67

Confidence intervals for the percentiles of non-durable consumption in MT

Chart 68

Confidence intervals for the percentiles of non-durable consumption in SK
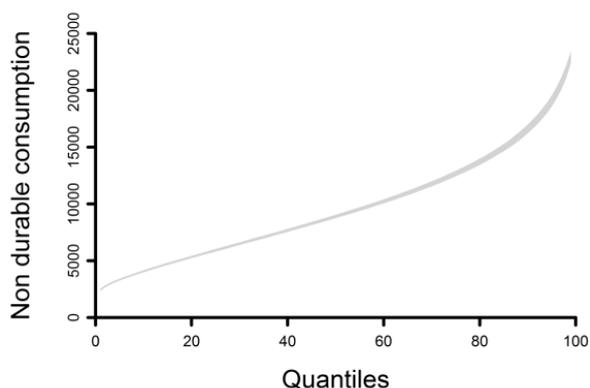
The variation for one given household may be quite strong, as the estimation at the individual level suffers from considerable uncertainty. However, the entire estimation is estimated in a very stable way. For instance, median consumption remains very stable across the different implicates: the coefficient of variation related to this estimator is about 1.4% for BE, 1.3% for CY, 0.5% for ES, 0.4% for FR, 0.8% for FI, 0.4% for IT, 1.1% for LU, 1.0% for MT and 0.7% for SK. Thanks to these 1,000 implicates, it is always possible to assess the robustness of one given estimation.

Chart 69, **Chart 70** and **Chart 71** present a very rough estimation of the link between consumption and wealth thanks to a LOESS regression, without any regard to the other observable characteristics (in particular income or age) that may have an impact on such an estimation. The dark blue line represents the computation performed with the estimation of non-durable consumption in the HFCS; the grey area stands for the confidence interval as given by the 1,000 implicates computed for the Monte-Carlo exercise. It then provides an idea of where the estimation of such a link (especially with regards to the MPC) may suffer from greater uncertainty. As expected, the top of the wealth distribution remains the most sensitive one, as the population used for estimating the MPC at the top is rather small. Also the light blue line represents the link between wealth and non-durable consumption when the variable has been collected among households through a 'one-shot' question.

Chart 69

Link between consumption and gross wealth in ES

Sources: HFCS, author's computations. Link estimated with a LOESS regression, confidence intervals at 95%.

Chart 70

Link between consumption and gross wealth in FR

Sources: HFCS, author's computations. Link estimated with a LOESS regression, confidence intervals at 95%.

Chart 71

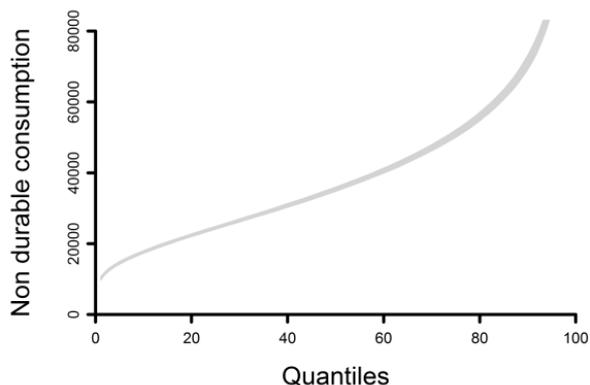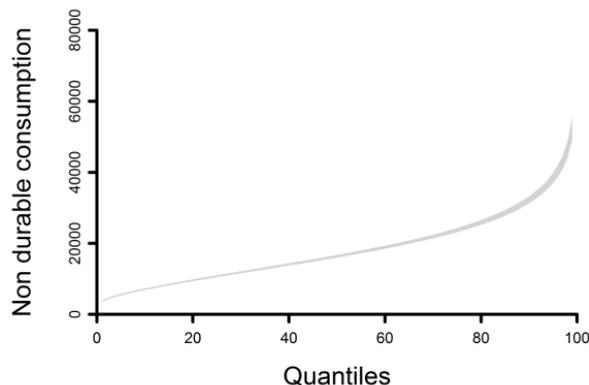Confidence intervals for the percentiles of non-durable consumption in IT

Sources: HFCS, author's computations. Link estimated with a LOESS regression, confidence intervals at 95%.
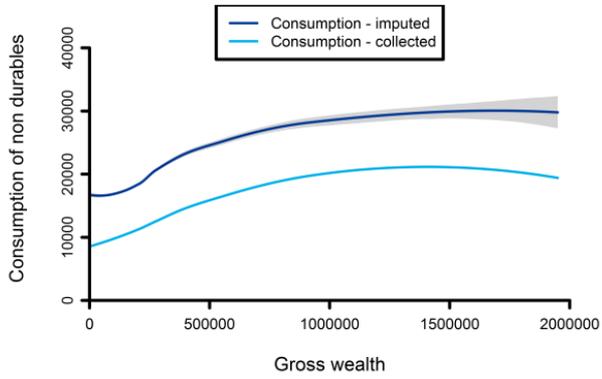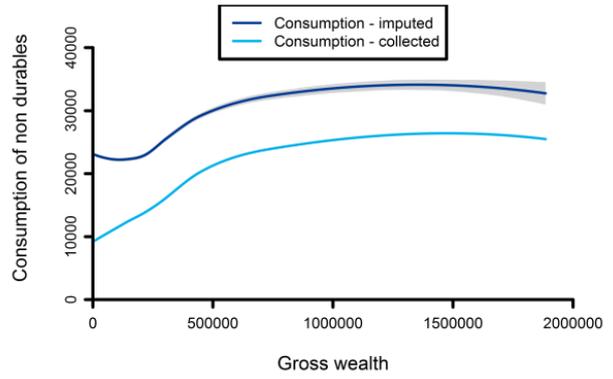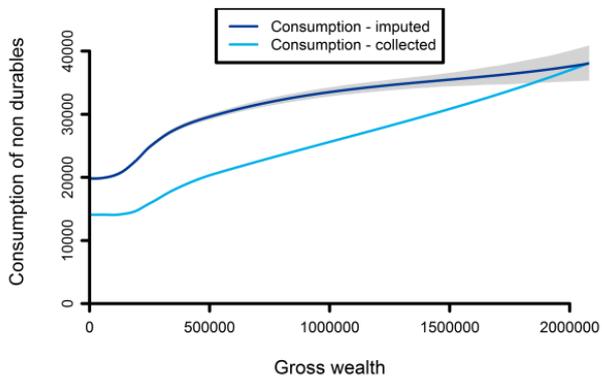
4.4.2 Second approach: trying to relax the Conditional Independence Assumption

Another source of uncertainty, as explained by D'Orazio et al. (2006), is related to the assumption that conditionally to the covariates, consumption does not depend on wealth or indebtedness for instance. This means basically that the entire correlation between consumption and wealth is carried out by the correlation between consumption, food consumption, age of the reference person and the other covariates that are used in the equation. This is questionable, as we would specifically like in the end to evaluate such a relationship.

The different conclusions that may be drawn from the analysis performed on the HFCS data may be assessed thanks to robustness checks consisting of relaxing the CIA assumption. Such an exercise implies then that the point estimate given by the main estimation would be compared with results provided by a range of plausible values that consumption may take. Conclusions that would be confirmed by such an exercise would then be highly reliable. It is nevertheless difficult to provide the user with such data. One simplistic way of estimating the reliability of the estimation would consist of using algorithms that aim to generate randomly conditional links between consumption and wealth. The algorithms enabling such a robustness check at the micro-level are not as simple as the Monte-Carlo exercise developed previously. Rässler (2004) describes the so-called NIBAS algorithm that enables to provide an idea of the range for the plausible values for consumption, using the fact that the variance-covariance between consumption and wealth has to be positive definite for defining the plausible values. This algorithm relies on assumptions about such a link, and for instance disregards to some extent the non-linearity that may occur at the top of the distribution. From this point of view, it may be seen only as a rough estimation, but for sure cannot so easily be considered as a robustness check per se.

The exercise carried out on the data shows unambiguously that the estimation of the distribution is much less stable once the CIA is relaxed. As shown in **Chart 72**, the grey area is wider than in **Chart 64**, partly also because the last figure does not represent a formal confidence interval, rather a range of plausible values. In that sense, the algorithm makes it possible to relax a considerable amount of hypotheses regarding the errors made when estimating consumption.

As a consequence, the range of possible values taken by the MPCs turns out to be much wider than before. The same exercise as the one shown in **Chart 70** is presented on **Chart 73**. The range of plausible values turns out to be far wider than was previously the case.

Chart 72

Range of plausible values for the distribution of consumption according to the NIBAS algorithm in FR
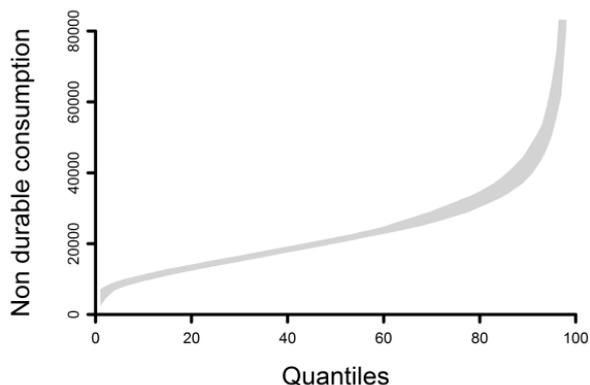
Sources: HFCS, author's computations.

Chart 73

Range of plausible values for the link between consumption and gross wealth in FR
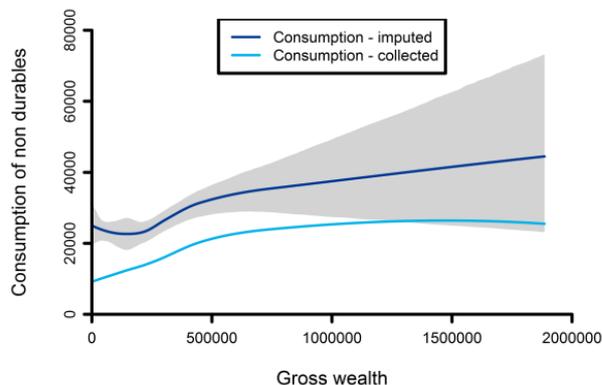
Sources: HFCS, author's computations. Link estimated with a LOESS regression.

4.5   Conclusions and caveats

The different estimations that have been tested over different countries in the HFCS make it possible to draw different conclusions for the sake of the estimation of consumption in a survey such as the HFCS. First it is essential to have at one's disposal variables that not only explain consumption very well, but also that are measured in a pretty consistent way between the HBS and the HFCS. Strong discrepancies in terms of distribution and/or totals may jeopardise the final results obtained through an estimation of consumption; in particular, the items of consumption (food consumption, utilities, and so on) have to be similar across both surveys.

Also it is important to keep in mind that the entire exercise relies on the Conditional Independence Assumption (CIA); from this perspective, the different analyses, in particular the ones aimed at describing the link between consumption and the variables that are not collected in the HBS may be conditioned by this initial assumption. It is then also interesting to provide the users with tools that will enable them to estimate the sensitivity of their results to this assumption.

As shown previously, the range of plausible values may be wide, once the CIA is given up. It would be also interesting to look for different ways to reduce the possible correlations between the different covariates. One way could be to add more variables into the NIBAS algorithm; this would probably reduce the interval of plausible values, but would also be far more demanding in computational terms for an already very cumbersome algorithm.

Finally, one should also be particularly cautious when using these estimated data as a dependent variable for a regression: as emphasised by Crossley et al. (2016), OLS procedure involving a dependent variable that has been estimated is biased: the

parameters are under-estimated, but can be corrected thanks to the R² of the OLS used for the estimation step (please refer to **Table 2**, last row for the correcting coefficients). Consumption may be computed with the method described above on the HFCS micro-data thanks to the scripts given in annex of this paper.

5 Alternative: the German and the Italian experiences

5.1 The German experience: collecting information on savings and building consumption as a residual

One experiment has been conducted with the first wave of the German part of the HFCS, aimed at collecting savings information directly throughout the entire questionnaire. The survey collects, for instance, information on regular saving flows allocated to each type of asset, but also discretionary savings and even dissaving. Net income is also computed thanks to a fiscal simulator, and information on debt service is also gathered. Hence the survey combines very precise pieces of information both on net income and net saving, which in the end makes it possible to compute very easily total consumption as a residual between net income and net saving. The **Chart 74** shows an overview of how this method performs for reproducing information on consumption in the HFCS.

Chart 74

Range of plausible values for the distribution of consumption according to the NIBAS algorithm in FR
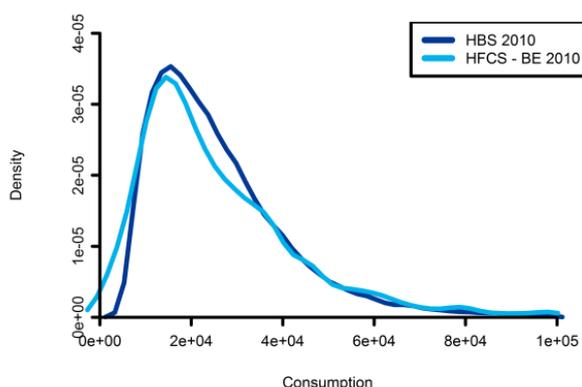
Sources: HFCS, author's computations.

The methodology and results of constructing total consumption from savings flows and income are described in Le Blanc and Schmidt (forthcoming). An exercise à la Skinner has also been conducted by the Bundesbank, following the same methodology as for the other countries. The results of such an exercise can be seen in **Table 2**, **Chart 27**, **Chart 44** and **Chart 53**.

5.2   The Italian experience

Most of the matching methods available in the literature are based on the conditional independence assumption (CIA) which assumes that the information shared between the two surveys (in particular socio-demographic variables) is sufficient to derive the joint distribution of interest. As mentioned previously, this assumption is questionable. Conti et al. (2015) propose an alternative method to perform statistical matching without CIA. The work uses data from two surveys: the first one is the Italian survey of household income and wealth (SHIW), the Italian part of the HFCS. The second survey is the Italian survey of households' budgets (HBS). As for the other presented techniques, the method assumes that expenditure is best estimated in HBS and rests on the questions about expenditure that have been asked in the SHIW. This information is used to construct bounds for the average propensities to consume of households with given demographic characteristics. A minimum value can be constructed using the information on expenditure coming from the SHIW survey (which is likely to be underestimated). On the other hand, a maximum value can be constructed using the HBS survey. In particular, the work identifies the class of all the joint distributions compatible with the marginal information on income and expenditure in the two surveys and which meet some constraints on the average propensities to consume. Within the set of all the plausible distributions, one joint distribution for household income and expenditure is selected via the Iterative Proportional Fitting algorithm. This joint distribution is then used to impute consumption for each SHIW respondent.

References

Aron, J. et al. (2012), "Credit, housing collateral, and consumption: evidence from Japan, the UK, and the US", Review of Income and Wealth, 58, No. 3, pp. 397-423.

Audenis, C., Grégoir, S. and Louvot, C. (2002), The various measures of the saving rate and their interpretation, OECD, Paris.

Baldini, M., Pacifico, D. and Termini, F. (2015), "Imputation of missing expenditure information in standard household income surveys", DEMB Working Paper Series, Modena.

Blundell, R., Pistaferi, L. and Preston, I. (2004), Imputing consumption in the PSID using food demand estimates from the CEX.

Browning, M., Crossley, T. F. and Weber, G. (2003), "Asking consumption questions in general purpose surveys", The Economic Journal, F540-F567 (113.491).

Carroll, C. D., Otsuka, M. & Slacalek, J. (2011), "How large are housing and financial wealth effects? A new approach", Journal of Money, Credit and Banking, 43.1, pp. 55-79.

Cifaldi, G. and Neri, A. (2013), "Asking income and consumption questions in the same survey: what are the risks?", Bank of Italy Temi di Discussione (Working Paper), Volume 908.

Conti, P. L., Marella, D. and Neri, A. (2015), "Statistical matching and uncertainty analysis in combining household income and expenditure data", Temi di Discussione, Issue 1018, Economic Research and International Relations Area, Bank of Italy.

Crossley, T., Levell, P. and Poupakis, S. (2016), Regression with an imputed dependent variable, mimeo.

D'Orazio, M., Zio, M. D. and Scanu, M. (2006), Statistical matching: Theory and practice, John Wiley & Sons.

Eurostat (2013), Statistical matching of EU-SILC and the Household Budget Survey to compare poverty estimates using income, expenditures and material deprivation, European Commission, Luxembourg.

Hamermesh, D. S. (1982), Consumption during retirement: The missing link in the life cycle.

HFCN (2013), "The Eurosystem Household Finance and Consumption Survey: Methodological Report from the First Wave", ECB Statistics Paper Series, Issue 1.

HFCN (2013), "The Eurosystem Household Finance and Consumption Survey: Results from the First Wave", ECB Statistics Paper Series, Issue 2.

Lamarche, P. (2015), "Can your stomach predict your total consumption? Results from the HFCS", Volume 39, Bank for International Settlements, Basel.

Le Blanc, J. and Schmidt, T. (forthcoming). Household consumption in Germany. New evidence from micro-data.

Lettau, M., Ludvigson, S. and Steindel, C. (2002), "Monetary policy transmission through the consumption-wealth channel", Economic Policy Review, Federal Reserve Bank of New York.

Lydon, R. (forthcoming), Household wealth and consumption: evidence for Ireland, Central Bank of Ireland Research Technical Paper.

Marcus, J., Sieger, R. and Grabka, M. M. (2013), Preparation of data from the new SOEP consumption module, DIW, Berlin.

Muellbauer, J. (2010), Household decisions, credit markets and the macroeconomy: implications for the design of central bank models, Bank for International Settlements, Basel.

Rässler, S. (2004), "Data fusion: identification problems, validity, and multiple imputation", Austrian Journal of Statistics, 33(1-2), pp. 153-171.

Rubin, D. B. (1986), "Statistical matching using file concatenation with adjusted weights and multiple imputations", Journal of Business & Economic Statistics, 4(1), pp. 87-94.

Skinner, J. (1987), "A superior measure of consumption from the panel study of income dynamics", Economic Letters, 23(2), pp. 213-216.

Slacalek, J. (2009), "What drives personal consumption? The role of housing and financial wealth", The BE Journal of Macroecomomics, 9(1).

Stiglitz, J., Sen, A. and Fitoussi, J.-P. (2009), The measurement of economic performance and social progress revisited, Reflections and overview, Commission on the Measurement of Economic Performance and Social Progress, Paris.

Sutherland, H., Taylor, R. & Gomulka, J. (2002), "Combining household income and expenditure data in policy simulations", Review of Income and Wealth, December, 48(4), pp. 517-536.

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

Appendices

Table 6
Classification of Individual Consumption by Purpose (COICOP)

Code	Items
CP01	Food and non-alcoholic beverages
CP02	Alcoholic beverages, tobacco and narcotics
CP03	Clothing and footwear
CP04	Housing, water, electricity, gas and other fuels
CP05	Furnishings, household equipment and routine maintenance of the house
CP06	Health
CP07	Transport
CP08	Communication
CP09	Recreation and culture
CP10	Education
CP11	Restaurants and hotels
CP12	Miscellaneous goods and services

Table 7
Differences in distribution of food consumption

	BE	CY	ES	FI	FR	IT	LU	MT	SK
Mean	27.8%	52.7%	-15.8%	n.a.	23.8%	-8.0%	45.1%	5.7%	25.1%
P5	20.2%	102.6%	30.8%	n.a.	40.7%	41.4%	43.2%	-2.9%	15.1%
P10	30.3%	67.7%	8.4%	n.a.	38.3%	4.7%	31.3%	-3.2%	13.6%
P25	31.1%	51.8%	-6.1%	n.a.	29.5%	0.8%	55.4%	3.8%	15.6%
P50	27.2%	45.3%	-7.3%	n.a.	12.0%	6.9%	41.2%	9.2%	30.8%
P75	17.0%	51.3%	-22.7%	n.a.	-9.7%	-14.3%	37.0%	7.2%	29.0%
P90	33.4%	36.7%	-24.0%	n.a.	-2.5%	-19.0%	45.0%	7.3%	33.3%
P95	34.1%	71.8%	-17.9%	n.a.	-1.5%	-16.1%	49.6%	3.3%	27.2%

Sources: HFCS and HBS. Authors computations.

Table 8
Differences in distribution of rents

	BE	CY	ES	FI	FR	IT	LU	MT	SK
Mean	22.4%	14.6%	5.7%	-5.3%	44.9%	-1.7%	2.2%	128.8%	12.8%
P5	79.0%	-	5.1%	-0.4%	300.0%	-13.1%	-21.5%	47.4%	160.4%
P10	39.0%	-	7.9%	-6.6%	314.8%	-21.8%	-25.8%	564.4%	50.0%
P25	23.5%	58.7%	14.1%	-5.3%	72.9%	-11.3%	-11.5%	276.5%	24.3%
P50	13.7%	4.2%	9.4%	-5.5%	37.4%	0.0%	-1.5%	154.5%	12.4%
P75	11.1%	1.3%	4.1%	-6.2%	20.0%	6.3%	0.8%	186.8%	10.0%
P90	12.0%	4.4%	6.8%	-5.1%	15.7%	1.6%	15.2%	260.0%	0.0%
P95	18.1%	-2.4%	7.0%	-3.4%	12.0%	0.0%	31.7%	63.6%	5.4%

Sources: HFCS and HBS. Authors computations.

Table 9

Differences in percentage point between the repartition as measured by HFCS and HBS

	BE	CY	ES	FI	FR	IT	LU	MT	SK
Age of RP									
20-	0.2%	-	0.0%	0.0%	0.0%	-0.1%	-0.1%	n.a.	0.0%
21-30	3.3%	3.1%	-1.2%	0.3%	-0.5%	-0.1%	1.0%	n.a.	3.5%
31-40	-0.1%	1.2%	0.9%	-0.0%	1.1%	-1.7%	-2.4%	n.a.	-0.9%
41-50	1.1%	0.0%	0.7%	1.3%	-1.5%	2.1%	2.6%	n.a.	1.7%
51-60	-4.1%	0.3%	0.1%	1.7%	-0.2%	-1.3%	0.1%	n.a.	-0.8%
61-70	-2.3%	-1.3%	0.0%	-2.0%	0.2%	-0.3%	-1.4%	n.a.	0.1%
70+	1.9%	-3.4%	-0.6%	-1.2%	0.9%	1.5%	0.2%	n.a.	-3.6%
Household size									
1	-0.3%	0.3%	-3.8%	-0.8%	0.5%	-5.0%	-1.1%	0.0%	2.7%
2	0.2%	3.5%	1.0%	-0.3%	-0.5%	3.3%	0.0%	0.0%	-1.8%
3	0.3%	0.6%	3.8%	0.5%	0.0%	-0.5%	0.9%	0.0%	0.7%
4	-0.5%	-2.4%	2.0%	0.3%	-0.7%	0.9%	-0.7%	0.0%	-1.0%
5+	-0.3%	-2.3%	-2.9%	0.3%	0.7%	1.4%	0.8%	-0.2%	-0.2%
Number of children									
0	-5.0%	-11.1%	-4.9%	-2.8%	0.0%	-0.9%	-6.3%	-15.6%	-12.4%
1	3.3%	3.1%	3.7%	1.1%	-0.2%	-1.1%	1.7%	4.4%	3.5%
2	1.0%	3.4%	1.9%	1.3%	-0.5%	0.8%	1.9%	5.8%	6.9%
3+	0.6%	4.4%	-0.7%	0.4%	0.8%	1.2%	2.6%	5.3%	1.8%
Level of education of RP									
1	1.3%	-7.4%	3.7%	-	3.2%	-0.7%	6.5%	n.a.	-10.0%
2	0.2%	-1.6%	-7.7%	2.1%	-22.4%	-7.2%	-13.4%	n.a.	-28.9%
3	-20.6%	-2.4%	-3.4%	5.3%	23.0%	8.3%	6.2%	n.a.	35.7%
5	20.8%	11.4%	7.4%	-8.7%	-3.8%	-0.3%	0.5%	n.a.	3.2%
Labour status of RP									
Employed	1.2%	1.8%	-3.8%	-0.6%	-1.4%	0.0%	2.4%	4.1%	13.2%
Unemployed	0.4%	0.0%	7.1%	5.7%	-0.4%	-6.5%	-3.6%	-6.9%	-7.7%
Retiree	0.5%	-1.7%	-3.3%	-5.1%	1.8%	6.4%	1.2%	2.7%	-5.4%
Tenure status									
Free use	-	-	0.2%	-0.8%	-	1.6%	-	-	0.2%
Home owner	0.9%	3.5%	2.4%	-0.6%	1.5%	-5.5%	-8.2%	12.6%	-2.5%
Tenant	-0.9%	-3.5%	-2.7%	1.5%	-1.5%	3.9%	8.2%	-12.6%	2.2%

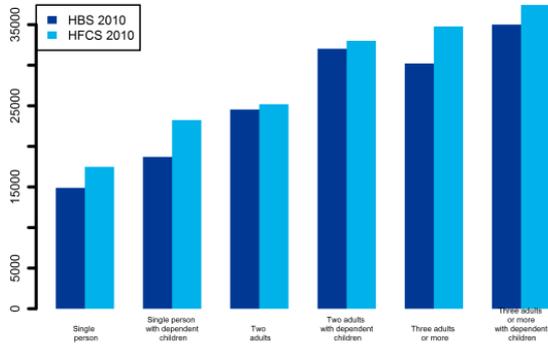Sources: HFCS and HBS. Authors computations.

Table 10

Ratio between the coefficients of variation of food consumption broken down by demographic variables

	BE	CY	ES	FR	IT	LU	MT	SK
Age of RP								
20-	0.88	-	0.85	0.57	0.49	-	n.a.	0.61
21-30	1.49	0.71	0.71	0.85	0.62	0.94	n.a.	1.15
31-40	1.11	0.76	0.74	1.98	0.71	1.01	n.a.	1.12
41-50	1.37	1.08	0.76	1.75	0.78	0.96	n.a.	0.98
51-60	0.97	0.97	0.72	1.54	0.81	0.91	n.a.	1.02
61-70	1.52	1.06	0.98	2.02	0.81	0.91	n.a.	1.19
70+	1.36	0.96	0.85	1.75	0.75	1.25	n.a.	1.11
Household size								
1	1.28	1.00	0.76	1.76	0.64	1.08	0.79	1.05
2	1.53	1.01	0.89	1.79	0.70	1.31	0.76	1.11
3	1.69	1.01	0.95	1.79	0.71	0.98	0.67	0.97
4	1.18	1.04	0.68	2.25	0.81	0.99	0.81	0.93
5+	1.44	0.97	0.79	1.63	0.90	0.91	0.86	0.81
Number of children								
0	1.27	0.96	0.83	1.70	0.73	1.11	0.91	1.10
1	1.69	1.19	0.62	1.89	0.72	1.02	0.95	0.92
2	1.09	0.93	0.94	2.05	0.81	1.03	0.85	0.98
3+	1.30	0.96	0.91	1.63	0.88	0.87	0.88	0.81
Level of education of RP								
1	1.26	1.12	0.87	1.96	0.71	1.02	n.a.	0.40
2	1.10	0.84	0.58	1.58	0.76	0.80	n.a.	1.43
3	1.50	0.92	0.81	1.80	0.78	1.00	n.a.	1.08
5	1.15	1.01	0.90	1.37	0.83	1.12	n.a.	1.06
Income quintile								
1	1.65	1.00	0.77	1.55	n.a.	1.08	0.92	1.31
2	1.86	1.00	0.73	1.91	n.a.	0.94	0.93	1.15
3	1.64	1.03	0.78	2.09	n.a.	0.94	0.89	1.14
4	1.22	0.82	0.54	2.16	n.a.	1.14	1.08	1.06
5	1.31	1.03	0.86	1.66	n.a.	1.21	0.94	0.93

Sources: HFCS and HBS. Authors computations. A ratio less than 1 indicates that the coefficient of variation in HBS is higher than the one in the HFCS.
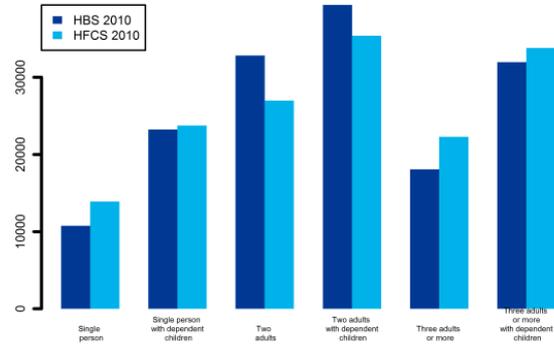
Chart 75

Comparison of medians by type of household in BE

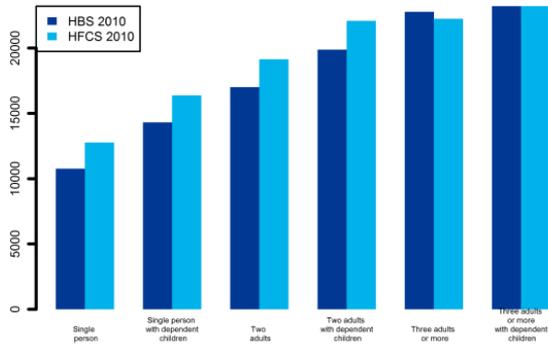Sources: HFCS and HBS, author's computations.

Chart 76

Comparison of medians by type of household in CY

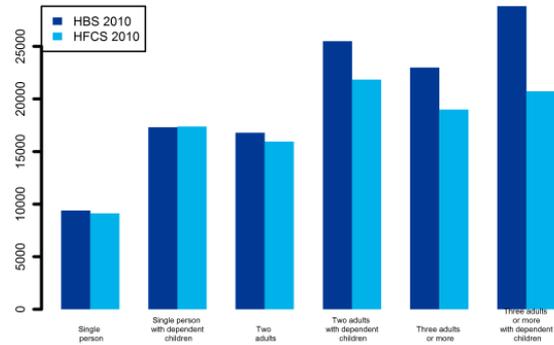Sources: HFCS and HBS, author's computations.

Chart 77

Comparison of medians by type of household in DE

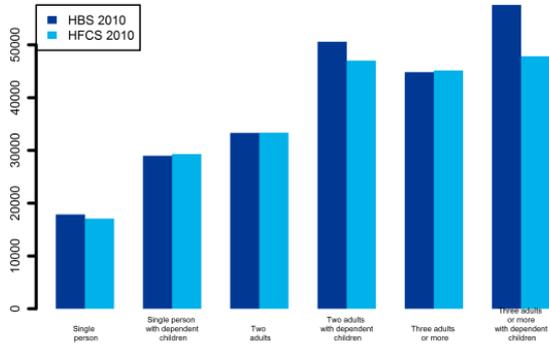Sources: HFCS and HBS, computations by Bundesbank.

Chart 78

Comparison of medians by type of household in ES

Sources: HFCS and HBS, author's computations.
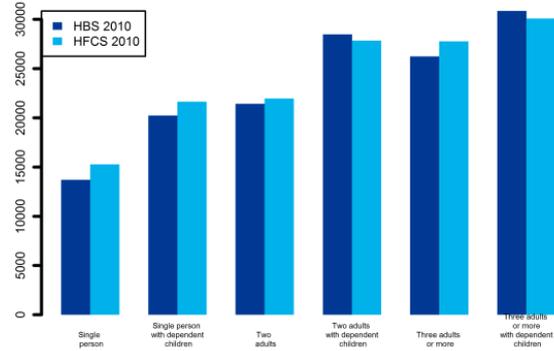
Chart 79

Comparison of medians by type of household in FI

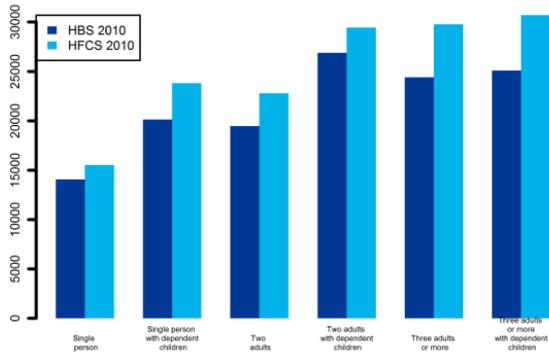Sources: HFCS and HBS, author's computations.

Chart 80

Comparison of medians by type of household in FR

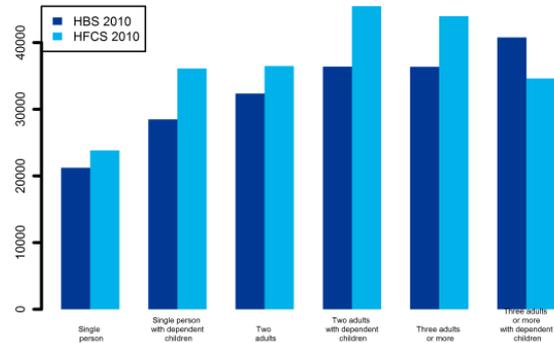Sources: HFCS and HBS, author's computations.

Chart 81

Comparison of medians by type of household in IT

Sources: HFCS and HBS, author's computations.
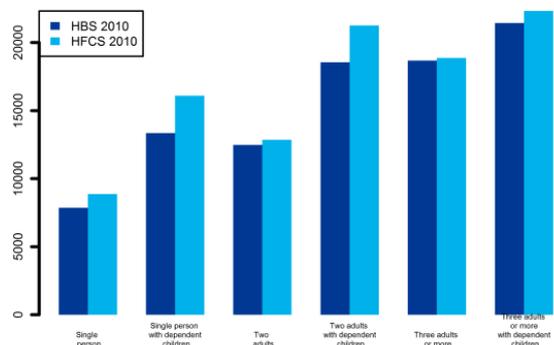
Chart 82

Comparison of medians by type of household in LU

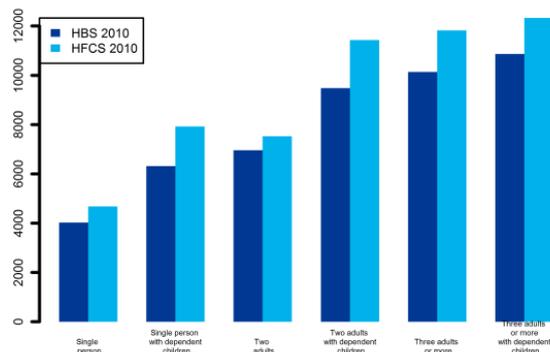Sources: HFCS and HBS, author's computations.

Chart 83

Comparison of medians by type of household in MT

Sources: HFCS and HBS, author's computations.

Chart 84

Comparison of medians by type of household in SK

Sources: HFCS and HBS, author's computations.

Table 11

Final specifications for the consumption equation

(OLS coefficients for the estimation of non-durable consumption in HBS data; stars on the left side of the figure indicate the statistical significance of the estimations: * 10%, ** 5". *** 1%)

Covariates	BE	CY	DE	ES	FI	FR	IT	LU	MT	PT	SK
Constant	8.98***	8.83***	17.83***	7.14***	9.44***	8.57***	0.24	2.17	11.98	8.48***	1.72***
Log food consumption at home	0.08	-0.54**	-4.36	0.24***	-	0.24***	0.11	2.65	-1.64	-0.13	2.30***
Log food consumption at home^2	-0.05	0.09*	0.74	-0.04**	-	-0.08***	-0.09	-0.35	0.18	-0.02	-0.32***
Log food consumption at home^3	0.01*	-0.00	-0.06	0.01***	-	0.01***	0.01	0.02	-0.00	0.01***	0.02***
Log food consumption away	0.09***	0.05	-	-	-	0.12***	-	0.06**	0.07**	0.12***	0.09***
Log food consumption away^2	-0.03***	-0.2*	-	-	-	-0.04***	-	-0.02***	-0.02**	-0.04***	-0.03***
Log food consumption away^3	0.00***	0.00***	-	-	-	0.00***	-	0.00***	0.00***	0.00***	0.00***
Log utilities	-	-	-	-	-	0.19***	-	-	-	-	-
Log utilities^2	-	-	-	-	-	-0.05***	-	-	-	-	-
Log utilities^3	-	-	-	-	-	0.00***	-	-	-	-	-
Log rent	-0.38*	-0.02	-33.86***	0.10***	-1.91*	0.12***	4.17	-0.16	0.08	-0.09	0.02
Log rent^2	0.07	-0.00	6.45***	-0.03***	0.41*	-0.03***	-0.66	0.01	-0.02*	0.01	-0.01
Log rent^3	-0.00	0.00	-0.54***	0.00***	-0.02	0.00***	0.04	0.00	3.26**	0.00	0.00*
Age of RP											
30-	-0.04*	-0.04	Ref.	0.02	0.28***	0.03***	0.01	-0.01	-	Ref.	-0.00
30-40	-0.03**	0.00	-0.01	0.00	0.25***	0.01	0.01	-0.01	-	0.00	-0.01
40-50	Ref.	Ref.	-0.05	Ref.	Ref.	Ref.	Ref.	Ref.	-	-0.02	Ref.
50-60	0.01	-0.04	-0.04	0.01	0.04	0.01	0.00	0.03	-	-0.03	-0.03**
60-70	0.02	-0.07**	-0.01	0.01	0.10	0.02	-0.04***	0.06*	-	-0.05	-0.01
70+	-0.02	-0.17***	0.07*	-0.08***	-0.15*	-0.03*	-0.13***	-0.04	-	-0.11***	-0.09***
Head male	-0.05***	-	-0.07***	-0.05***	0.02	-0.10*	-0.02***	-0.03**	-0.01	-0.03**	-

Size of household											
1 person	-0.07***	-0.13***	Ref.	-0.09***	-0.29***	-0.06***	-0.13***	-0.08***	-0.07***	Ref.	-0.07***
2 persons	Ref.	Ref.	0.04***	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	0.10***	Ref.
3+ persons	0.05***	0.05**	0.05***	0.02	-0.11	0.02	0.06***	-0.08***	0.05**	0.13***	0.05***
Tenure status											
Owner	0.07*	-0.02	Ref.	0.05*	0.41***	-0.13***	9.45	-0.07	0.01	-0.07**	0.10***
Tenant	Ref.	Ref.	65.15***	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Free use	-	-	-0.15***	0.02	0.49***	-	9.37	-	-	-0.06***	-0.03
# of children											
No child	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	-	Ref.	Ref.
1 child	0.00	0.06**	0.01**	0.04***	0.28	0.02	-0.01	0.00	-	0.03	-0.01
2 children	0.01	0.08***	0.02***	0.09***	-0.11	0.01	-0.04**	0.02	-	0.03	-0.02
3+ children	0.01	0.10***	0.02**	0.12***	0.40	-0.00	-0.04*	0.05	-	0.05	-0.05**
Level of education											
0+1	-0.08***	-0.10***	Ref.	-0.13***	-	-0.10***	-0.17***	-0.10***	-	-	-0.10***
2	-0.05***	-0.01	-	-0.06***	0.10***	-0.04***	-0.08***	-0.01	-	-	-0.05***
3	Ref.	Ref.	0.03***	Ref.	Ref.	Ref.	Ref.	Ref.	-	-	Ref.
5+6	0.05	0.02	0.07***	0.07***	0.11*	0.04***	0.10***	0.04**	-	-	0.05***
Labour status											
Employed	Ref.	Ref.	-	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Unemployed	-0.01	-0.06**	-	-0.05***	-0.35***	-0.02*	-0.11***	-0.05*	-0.08***	-0.02	-0.04***
Retiree	0.00	-0.00	-	-0.02	-0.18**	0.01	-0.03**	0.01	-0.13***	-0.01	-0.01
Income quintile											
1	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
2	-5.13	1.66	0.13***	1.61***	0.68***	0.09**	-0.69	8.00	13.56	0.17*	13.90
3	-0.19	-0.45	0.23***	11.71***	0.97***	0.24***	-5.28	6.23	-2.55	0.09	2.47
4	-0.39*	11.85***	0.33***	0.29	1.13***	0.41***	1.88	-1.95	-0.76	0.32***	29.68
5	0.32*	-2.14	0.49***	2.33***	1.21***	0.58***	0.37	10.65	60.00***	0.76***	9.96

Sources: HBS data. Computations by Bundesbank for DE, Banco de Portugal for PT, author's computations for the other countries.

Pierre Lamarche

At the time of writing, Pierre Lamarche was at the European Central Bank, Frankfurt, Germany; now at Eurostat, Luxembourg;
email: hfcs@ecb.europa.eu