Can your stomach predict your total consumption?

Results from HFCS data

Pierre Lamarche

Abstract

As consumption is one of the main reasons why households accumulate income and build up wealth, micro-data containing information about households' wealth, income, and consumption together would be an important asset. Such micro-data would enable analysis on many economic questions in general and for monetary policy in particular. For example, it would be possible to address the question of wealth effects at a micro-level taking into account the heterogeneous behaviours among the population, and, with information about indebtedness, to investigate the sensitivity of consumption to credit constraints.

Convinced by the relevance of such data, one would intend to implement questions on consumption in existing surveys on wealth and income. We follow the framework described by Browning, Crossley and Weber (2003) and use the questions on expenditures for food included in the Eurosystem Household Finance and Consumption to predict total consumption. An experiment has been carried out on the French part of the survey: food expenditure at home and outside is a good predictor for total consumption, explaining two thirds of the variation of consumption. The predicted consumption in the HFCS is then compared with both micro and macro data. Thanks to these different comparisons the validity of the method can be assessed, resulting in a valuable micro-data that combine information about income, wealth, and consumption.

Keywords: Consumption, Survey, Household

JEL classification: D120, D140, D310

Contents

Introduction

From an economic perspective, studying wealth accumulation, indebtedness or labour offer among households is strongly related to their consumption behaviours. Indeed, according to the consumer's theory, consumption is the final aim for households and as such helps in explaining saving behaviours, portfolio choices or indebtedness. To that respect, a micro-based analysis would in an ideal world use data containing information about wealth, income and consumption at the same time. More specifically, analyses for monetary purposes such as estimations of the size of wealth effect or the reaction of consumption to credit constraint would greatly benefit from micro-data combining such pieces of information.

The Eurosystem Household Finance and Consumption Survey (HFCS), conducted in the euro area by National Central Banks, aims at gathering information on these three aspects among households. This survey is very similar to the Survey of Consumer Finances (SCF) conducted by the Federal Reserve in the U.S. The questionnaire which is already quite long collects very precise pieces of information about assets held by households. It also describes accurately indebtedness, distinguishing mortgage and non-mortgage loans. Understanding the role of income in the process of wealth accumulation or for the access to finance is a key issue. This is the reason why the questionnaire includes a part dedicated to the income of the households, describing the different sources of income that the household benefit from and similar to the framework used in the EU-SILC for instance. Finally the questionnaire includes a few questions about consumption and in particular food consumption. These questions aimed at addressing the issue of consumption for households whose income and wealth are known.

To study accurately the link between consumption, income and wealth, one would be interested in having a very precise description of the consumption expenditures for each household included in the survey. However, in this context, the direct collection of consumption data is challenged by the length of the questionnaire. Indeed, as already emphasised by Browning, Crossley and Weber, there are different ways to collect data on consumption. The most exhaustive one (but also of course the most demanding one) is based on diaries in which households' expenditures are reported, say on a weekly basis. Regarding the duration of the questionnaire, costs of data collection and response burden for the households, such a solution appears hardly feasible in the context of the HFCS.

This is the reason why a second best solution was implemented in the HFCS questionnaire. Following the recommendations by Browning, Crossley and Weber (2003) for the estimation of consumption expenditures in general-purpose surveys, a limited number of questions on aggregate components of consumption have been included in the HFCS questionnaire. Of course these questions do not describe the entire consumption; however, the framework described by Browning et al. should enable to estimate the expenditures for total consumption. Indeed, the methodology they implement uses the information in surveys such as the Household Budget Survey (HBS) so to reproduce the link between, say food consumption and total consumption as measured in consumption-oriented surveys.

Literature

Collecting information about consumption is a question that has already been investigated in the literature. As a start, Hamermesh (1984) compares the measure of consumption obtained thanks to the Retirement History Survey (RHS) with those given by the Consumer Expenditure Survey (CEX). The Retirement History Survey aims at following retired people; the CEX is the official survey for measuring consumption expenditures. Investigating for a possible decrease of the consumption due to retirement, the author takes advantage of the longitudinal data and the few questions about consumption provided by the RHS to assess this potential decrease. Indeed, one of the main drawbacks of the standard consumption surveys consists of their lack of longitudinal information. For instance households followed by the CEX are included in the survey for only one year at most, which is insufficient to assess a change through time due, say to retirement. To ensure the validity of information about consumption in the RHS, Hamermesh compares the ratio between consumption and income measured in both surveys. Taking advantage of this result, Skinner (1987) uses information about the link between food consumption, rents and value of the house estimated on the CEX to compute total consumption in the Panel Study of Income Dynamics (PSID). The method implemented in this article succeeds in explaining up to 78% of the variance of the consumption in the CEX, providing a simple method to estimate total consumption in such surveys.

This framework has been then used in different surveys, using various methodologies and investigating multiple questions. For instance, Blundell, Pistaferri and Preston (2004) implemented a more sophisticated method on the CEX, always to impute consumption in the PSID. They estimate a standard demand function for food, which can be seen as an inversion of the equation estimated by Skinner. They address the issue of potential non-linearity in such demand functions and take for potential endogeneity through time of total consumption component. Browning, Crossley and Weber (2003) provide a general overview of the different experiments that have been conducted since. They recall the importance of the measurement of consumption in general purpose surveys and provide a list of topics that could significantly benefit from such improvements. Moreover they describe and compare the many different ways of collecting information about consumption in surveys that would greatly benefit from such an improvement but that are not initially intended to gather such pieces of information. They first recall the results of experiments consisting of the inclusion of a question on broad consumption (such as "How much do you spend on everything in a typical month?") that have occurred in surveys such as the Canadian Out of Employment (COEP) or the Italian Survey on Household Income and Wealth (SHIW). They find that total consumption tends to be strongly underreported by households with only one question. This underreporting phenomenon has been confirmed by other experiments (see for example Cifaldi and Neri, 2013). However, since bias due to underreporting remains consistent over countries and experiments, they suggest the existence of methods to correct for underreporting in the estimation process.

An alternative way of collecting information about consumption consists of asking questions on very precise and limited sub-items. One typical candidate is food consumption; indeed this kind of consumption appears to be easily identified by households and seems to suffer from a less important underreporting bias.

Experiments that have been conducted on the CEX or the SHIW are quite conclusive, according to Browning et al. The only major concern about the inclusion of this sole question in the survey is that it brings sometimes not enough information to properly estimate total consumption. The best way to collect information is to submit an exhaustive list of sub-items to the household. However this kind of solution is quite demanding for the respondent and does not differ so much compared to a diary survey in terms of response burden. The best trade-off between response burden and data accuracy is then a non-exhaustive list of sub-items. Thus, not only questions on food at home consumption can be included in the questionnaire, but also questions on utilities or food at restaurant. Then applying the method given by Skinner with data from a standard consumption survey should enable to estimate total consumption in general purpose surveys.

Data and methodology

The first wave of the Eurosystem Household Finance and Consumption Survey has been conducted between 2010 and 2011 in most countries. This survey is conducted in a decentralised way. Each institution participating to the Network (National Central banks or National Statistical Institute) is responsible for conducting the survey within its country, and the European Central Bank is coordinating the Network toward a common methodology. Finally, about 63,000 households living in the euro area have participated in this survey (see table 1). Information about demographics, financial and real assets, financing for the purchase of this assets, liabilities, credit constraints, employment, pensions, income, intergenerational transfers, and consumption are gathered through this survey among these households.

Sample size in the HFCS

Table 1

Country	Net sample size
Belgium	2,364
Germany	3,565
Greece	2,971
Spain	6,197
France	15,006
Italy	7,951
Cyprus	1,237
Luxembourg	950
Malta	843
Netherlands	1,301
Austria	2,380
Portugal	4,404
Slovenia	343
Slovakia	2,057
Finland	10,989

Source: HFCN (2013)

One of the main features of the survey is that the countries participating in the project follow an ex-ante harmonised methodology. In particular, the variables that are included in the survey have been elaborated according to a common set of definitions and concepts. However, even if a common blueprint questionnaire has been elaborated and is the starting point, national questionnaires may differ substantially for two main reasons. First, for some countries such as France, Italy or Spain, the survey pre-existed to the HFCS. Indeed, the Enquête Patrimoine in France exists since 1986 and the Spanish Encuesta Financiera de las Familias since 2002 and even the Italian Survey on Household Income and Wealth (SHIW) since the 1960s. For these countries, there was a need for questionnaire convergence and a few variables have been adapted from the original survey variables. Second, national specificities may imply some adaptations of the questionnaire, in particular in the wording of the questions and the use of national concepts, so to ease the comprehension of the questionnaire by the respondents.

Following the framework by Skinner (1987), we use the data from the Household Budget Survey (HBS) conducted by Eurostat, which provide a precise description of the yearly consumption for the households. This survey has been conducted within the member states of the European Union in 1988, 1995, 1999, 2005 and for its last wave, in 2010. The coincidence of the HBS last wave with the first wave of the HFCS is of course highly valuable in the perspective of reproducing Skinner's methodology on those data. As emphasised by Eurostat, the main purpose of the HBS is to collect information that will allow updating the weights of the bucket of goods and services used in the computation of the Harmonised Index of Consumer Prices. However, the data can be used or other purposes related to consumption. The data collects information about household consumption using the Classification Of Individual Consumption by Purpose (COICOP). This classification enables to identify in the HBS data for instance what can be considered as food consumption (see table 2).

Classification of Individual Consumption by Purpose

Table 2

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	Heath
CP07	Transport
CP08	Communications
CP09	Recreation and culture
CP10	Education
CP11	Restaurant and hotels
CP12	Miscellaneous goods and services

Source: Eurostat

One of the main features of these two surveys is the fact that they both result from gentlemen agreements and are not enforced by any European regulation. The demanding harmonisation work is then conducted by the different European networks in charge of the two surveys. This could also lead, in our perspective of applying a general methodology on these data, to the choice of adapting the method to the specificities of the national data we are facing.

Basically, Skinner's method can be justified by the idea of matching based on observations shared by both surveys. Moreover, Browning, Crossley and Weber offer a theoretical framework to justify the method. Conversely to Skinner, they focus only on non-durable consumption that seems to be more correlated to food consumption than total consumption. We follow this framework and offer a solution for estimating durable consumption. We then have a list of sub-items $(c_1 \dots c_n)$ and for each of them, one question has been asked in the HFCS. We focus on three particular components: food at home, food outside home, and utilities (water expenditures, electricity, fuel and communications), and link these components with total expenditures for non-durables using a linear Engel curve specification:

$$x_i = \alpha_i + \beta_i x + u_i$$

where x and x_i are respectively the total expenditure for non-durables and the expenditures for item i. Then, taking into account for instance the weights ω_i of each item i in non-durables consumption, Browning, Crossley and Weber obtain the estimating equation:

$$x = \left(-\sum_{j=1}^{J} \alpha_j \frac{\omega_j}{\beta_j} \right) + \frac{\omega_1}{\beta_1} x_1 + \cdots + \frac{\omega_J}{\beta_J} x_J - \sum_{j=1}^{J} \frac{\omega_j}{\beta_j} u_j$$

This equation can be estimated thanks to the HBS. The estimated coefficients are then applied on the target survey (e.g. the HFCS). In order to improve the model, Browning, Crossley and Weber also add some demographic covariates to their equation. Following the work by Blundell, Pistaferri and Preston, we use a log specification and we take into account potential non-linearity: for each sub-item, we use a polynomial specification.

Once the equation estimated, we can compute for each household belonging to the HFCS an estimation of their yearly expenditures for consumption of non-durables. To do so, we take the estimated parameters and apply them to the set of covariates used in the equation. For the unexplained part (the residuals), we face different options:

- The first one consists simply of taking the expectancy conditionally to the observables. As we have modelled the log of consumption, we make the assumption that the residuals follow a normal law $N(0,\sigma^2)$ and then:

$$E(C|X) = \log(\beta X)\, e^{\frac{\sigma^2}{2}}$$

- For the second one, we draw residuals from a truncated normal law, with the lower bound defined as the sum of the known components of the consumption (food consumption, utilities, rents). We also can take into account the total consumption of non-durables as declared by the household, when the question has been asked.

- For the last one, we use a method of stratified hot-deck over the residuals obtained through the HBS data, so to address potential heteroskedasticity of the residuals. This method has for example been applied by Cifaldi and Neri (2013). We can as well reproduce this hot-deck procedure say 1,000 times and compute an estimation of the expectancy for each household. Such a procedure could be viewed as a generalised residual method.

Once the consumption for non-durables estimated, we are naturally interested in having the total consumption. Thus we impute the consumption for non-durables, using the global consumption provided by National Accounts. This flow of expenditures denoted as F has to be allocated between the households. This is achieved by using additional information provided by the HFCS data about the stock of durables possessed by the households. Let s_i stand for the share of the household i over the global amount of durables. This share can easily be estimated with the HFCS data and we obtain the consumption for durables F_i of the household i:

$$F_i = F.s_i$$

Implicitly we assume that the depreciation rate δ is the same for every household. This is very unlikely to be true; however this assumption allows us to take into account the quantity of durables possessed by the household, which in a way reflects its consumption of durables.

Results

Response pattern

As previously underlined, the surveys used here are conducted on a voluntary basis. At the time being, the HBS data are not yet available for the whole euro area. We then conducted a first estimation on a few set of countries for which micro-data are already disseminated, namely Spain, Italy and France. For these countries, surveys on wealth and income pre-existed to the HFCS, and some effort to make the different questionnaire had to be made. In particular, this can explain the fact that all variables have not been collected in every country. This could affect the process of estimation and models have to be carefully adapted country by country. In particular, for Italy and Spain, food consumption at home and outside are combined in one single variable. Both countries did not collected expenditures for utilities, which was not classified as a core variable for the survey at the first wave. For France, things are slightly different: questions about expenditures for food consumption at home, outside and utilities were included in the questionnaire alongside with qualitative questions on the habits of consumption1. However these questions were located in a specific module which was applied to only one third of the sample. In that case weights used to compute the estimators have to be multiplied by 3, taking into account that selection for this module is random.

[1] These questions were focusing on the existence of regular expenditures for clothing, public transport, health, and so on.

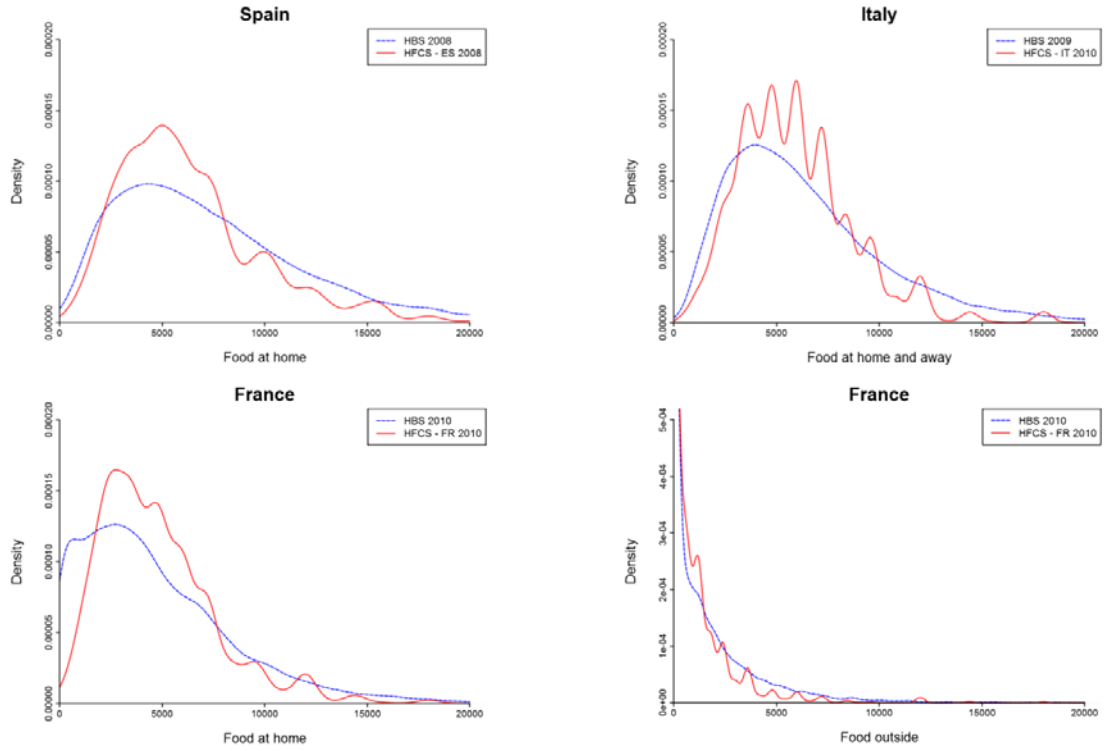
Response behaviours to the questions about specific sub-items of consumption do not reveal any difficulty or reluctance for households to answer. Indeed, about 98% of the Spanish and 100% of the Italian households answered to the question on food consumption. In France, the module on consumption was located at the end of the questionnaire, and despite a quite long interview, the response rate was quite good, although lower than Spain and Italy: indeed 89% of the households selected for this module accepted to answer to these questions. Finally, we have information about consumption for 6,064 Spanish, 7,951 Italian and 4,519 French households.

Comparison of covariates

A close look is then brought to the distribution of the covariates. Indeed, having a chance of getting a consistent estimation of consumption in the HFCS implies that covariates are distributed in the same way in both HFCS and HBS. In particular, Blundell, Pistaferri and Preston insist on the crucial need for similar data.

Comparison of the distributions of food consumption measured in the HFCS and in the HBS

Figure 1



Source: HFCS and HBS – author computations. Kernel density estimates; optimal Gaussian kernel.

For food consumption, despite completely different modes of data collection (in HBS information is collected through diaries when it is gathered thanks to one question in HFCS), the comparison between the two sources of distributions of food consumption shows quite consistent results. Indeed the global shape of the

distribution is preserved in the HFCS, and the accumulation points observed in the HFCS data are due to rounding by households. In Spain and Italy, the results are comparable especially in the bottom of the distribution. Thus the Italian 10th percentile in HFCS is 5% above the one observed in HBS (for more detailed results see annex 2) and the Spanish one is 8% above. However, for these two countries, analysis on the top of distribution gives worse results: the 90th percentile is respectively 24% and 19% below expected in Spain and Italy. For France, things are quite different, since the problem is located at the bottom of the distribution. Indeed, the 90th percentile in HFCS data is 3% below the one measured in HBS data, but the 10th percentile is highly overestimated (+38%). However, this could be explained by the fact that fieldwork has been shortened in France between the 2005 wave and the 2010 wave. In 2005, households were asked to fill diaries for two weeks, whereas they had to do so for only one week in 2010. This change in data collection could have an unexpected effect on the shape of the distribution, in particular for households in the bottom of the distribution that are more likely to go to the supermarket only once or twice a month. For those households, the shape of the distribution could be poorly estimated. Indeed experiments on HBS 2005 data show far better results for France.

To include income or not

Despite these results, we chose to use the 2010 wave for France for different reasons. First, we want to use comparable data for the consumption estimation in the different countries, in particular in terms of vintage. Second, we include in the equation not only food consumption but also a wide range of demographic variables that help in improving the quality of the model. In particular, the inclusion of income variables in the model should improve its quality. However there are pros and cons for including such variables in the equation. Browning, Crossley and Weber insist on the fact that they exclude income from the equation in spite of its predictive quality. They point out the fact that income could be poorly measured in surveys such as HBS. However, which appears important in such a work is that income is measured in the same way in both surveys. They also emphasise the potential spurious relationships that income could introduce in the equation. In order to take into account these limitations, we chose to use not income per se, but rather the deciles of income. Thus we introduce decile fixed effects and interactions so to allow coefficients to vary according to the decile of income.

Following results presented by Browning et al. and Blundell et al., we introduce in the equation demographic covariates: age and gender of the reference person , size of the household, number of children living in the household, tenure status and diploma of the reference person. These demographic variables should enable to take into account consumption behaviours that affect less food consumption. In particular, the age of the reference person carry away information about the position of the household in the life cycle. From the point of view of total consumption, the life cycle theory may not apply to the food consumption, whereas it applies to total consumption. Diploma of the reference person is commonly used as a proxy for permanent income. Here again, taking into account such pieces of information may have a value-added on the estimation of the equation. Indeed the estimation of the equation shows that such covariates have a significant effect on

the total consumption even when controlling from food consumption. It also increases the explanatory power of the model.

Explanatory power of the models

Table 3

Models – Adjusted R^2	Spain	France	Italy
Only food consumption	0.66	0.61	0.51
Food consumption, demographics	0.73	0.71	0.60
Food consumption, demographics, income dummies	0.75	0.77	-
Food consumption, demographics, income dummies and interaction	0.76	0.78	-

Adjusted R^2 obtained for models estimated on the HBS data

Indeed, as shown in table 3, food consumption explains already between the two-third and the half of the variation of total consumption in the HBS data. Adding demographics in the model has a positive impact on the explanatory power of the model, increasing the explanatory power up to 10 points. We test different specifications for the inclusion of such covariates in the model. First, following specifications shown by Browning et al., we include age of the reference person and household size with polynomial terms, which turns to be a demanding specification in terms of assumptions. The second approach consists of the inclusion of only dummy variables for demographics, without any change on the explanatory power of the model.

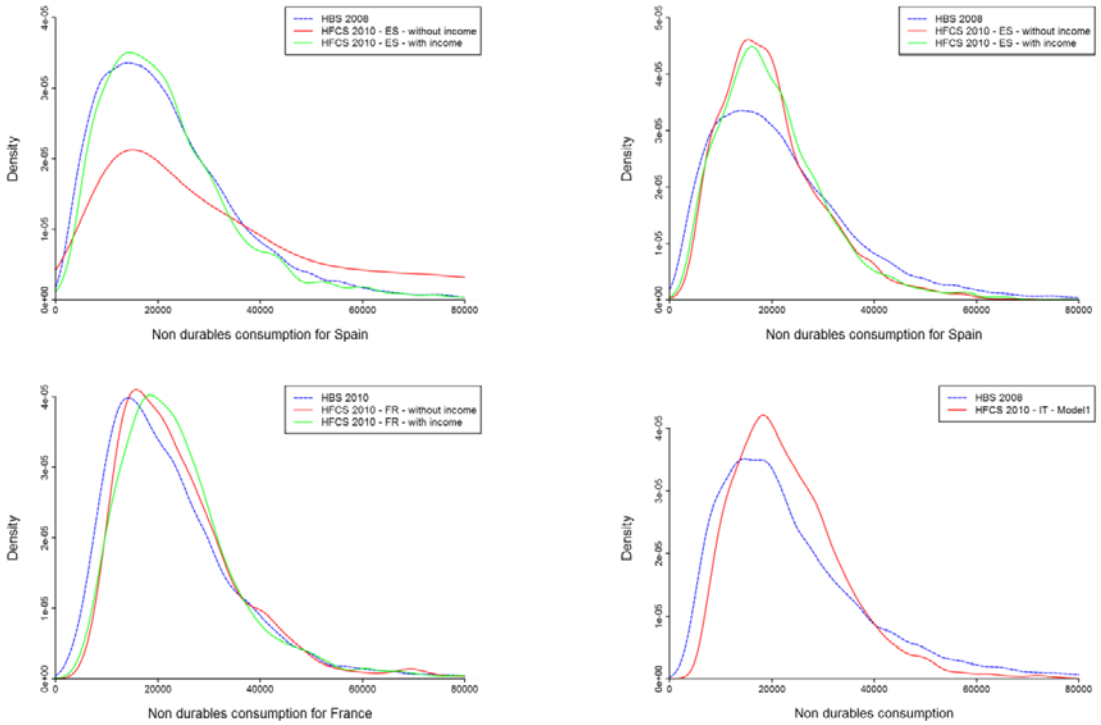
Including income has a quite limited effect on the improvement of the model in terms of predictive quality. However, analysis of the obtained distributions in the HFCS data shows that the distribution of consumption for non-durables obtained with the inclusion of income fits better the expected distribution (see figure 2), in particular for Spain. Furthermore, as shown for Spain, the first specification with polynomial terms for age and household size (first row of figure 2, left) gives far better results than the specification with dummy variables (first row of figure 2, right) in terms of fit with the expected distribution.

Moreover, we can compare the different choices for collecting information about consumption and determine the optimal way for such an exercise. In particular, some countries have collected food consumption at home and outside as a unique variable, while others have distinguished the two concepts. A first assessment on French data shows that even if the distinction of food consumption at home and away has not a huge impact on some indicators such as R^2, it still

improves the explanatory power of the model: a Chow test proves without ambiguity that the two models are not equivalent in terms of explained variation.

Comparison of the distributions of consumption for non-durables measured in in the HBS and estimated in the HFCS

Figure 2



Source: HFCS and HBS – author computations. Kernel density estimates; optimal Gaussian kernel.

Assessment of the effect of the residuals

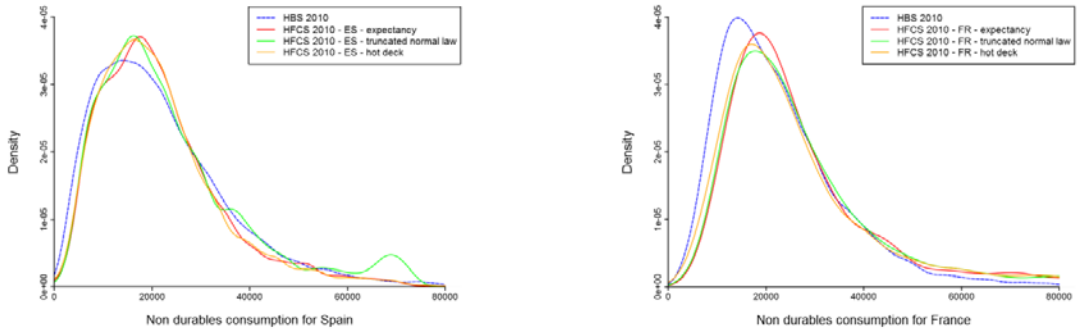
We assess the differences between the different ways of generating residuals when estimating consumption on the HFCS data. The first way of taking into account the unexplained part of the equation is to compute the expectancy for each household given its food consumption and other covariates that take place in the equation. The main drawback of such a solution is that it would potentially create points of accumulation in the distribution of consumption, since those households who share the same amount of food consumption and other characteristics will be imputed the same consumption. Moreover it does not enable to take into account some obvious bounds to the estimation such as the sum of sub-items of consumption that we have at our disposal in the data (food consumption, rents, utilities).

To take into account these bounds, we generate a truncated normal law with the lower bound corresponding to the maximum between total consumption as declared by the household (if the question has been asked) and the sum of the sub-items we have in the data. To address the issue of potential heteroskedasticity of the residuals, we finally use a stratified hot-deck procedure to allocate residuals

estimated on the HBS data to the HFCS data, following a stratification with the percentiles of food consumption.

Distribution of consumption for non-durables with the different methods for the residuals

Figure 3



Source: HFCS and HBS – author computations. Kernel density estimates; optimal Gaussian kernel.

As shown on figure 3, one method does not appear clearly preferable to the other ones from the point of view of distribution. There are very slight differences between the results. In particular, taking into account the bounds in the estimation does not change completely the result, which is also encouraging for the quality of the equation. However, the upper bound appears to have an impact on the top of the distribution for Spain. This is due to the polynomial specification, which leads to the introduction of outliers in the estimation. Indeed such a specification is very demanding in terms of assumptions.

Assessment of the accuracy

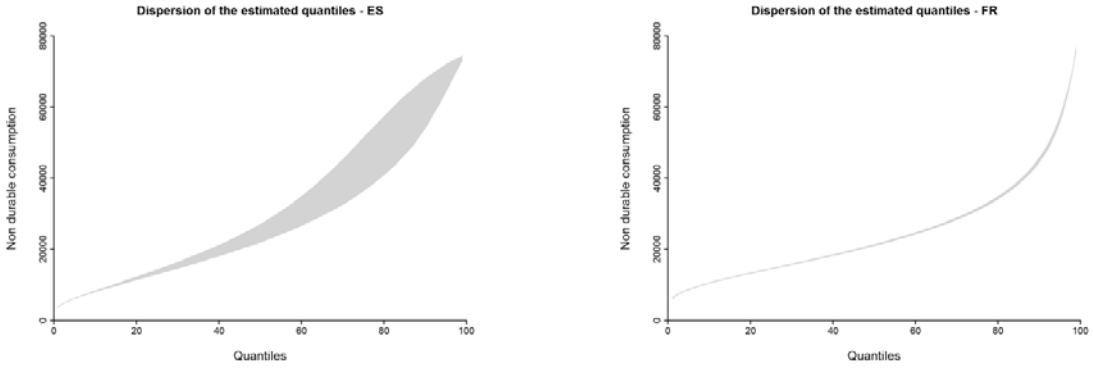
We then need to assess the uncertainty related to the estimation. As consumption for non-durables is not directly observed but estimated, we want to have an idea of the accuracy of this estimation. This can been done by simulating a high number of estimations taking into account the uncertainty due to the unexplained part of the equation, but also the uncertainty related to the estimation of the coefficients. To do so, we use the fact that the estimation of the parameters follows a normal law whose parameters can also be estimated. We generate 1,000 sets of coefficients and residuals according to their respective distribution and we obtain 1,000 estimations of total consumption.

The results of the simulations show with no ambiguity that there is a trade-off between a good fit with the expected distribution and the parsimony in the modelling. For instance, equations involving interactions between income and other covariates and polynomial terms for age and size result in high standard errors associated to the estimation of the coefficients. Then the simulation of these coefficients following those standard errors becomes pretty unstable and produces highly volatile estimations, with huge coefficients of variation. The use of a lower number of parameters (see annex 4 for a complete description of the models) has a

positive effect on the variance of the estimation, but as a result the fit with the expected distribution is deteriorated, in particular for Spain. As shown in figure 4, the variation of the estimated quantiles of consumption for non-durables remains pretty high for Spain. The dummy specification for France and Italy leads to a very more precise estimation. Thus, the median for France is estimated with a coefficient of variation of 0.4%; it reaches 5.4% in Spain. Results for Italy are very close to that obtained on French data. This uncertainty is represented in the figure with the grey area, which corresponds to the area where 95% of the estimations are located. From that point of view, comparison of the outcomes for the different specifications is pretty conclusive.

Confidence intervals for consumption percentiles

Figure 4



Source: HFCS - author computations.

Conclusion

The difficulties for estimating consumption in a survey like the HFCS can appear as high as the benefit from having such a variable at one's disposal in this survey. More precisely, defining a common approach for every country belonging to the euro area proves to be quite difficult: more detailed information about consumption collected by some countries has to be used in the model so to improve the outcome of the estimation. Close attention has to be put on the modelling process for which the specification may have huge effects both on the expectancy and the accuracy of the estimation. To that respect, assessment of the results has to take into account not only the comparison with the distribution in the consumption survey, but also the precision of the estimation. Moreover, results on Spain, France and Italy suggest that there is not one unique model that would predict with high accuracy total consumption. On the contrary, models have to be adjusted to match national specificities and reproduce as much as possible the expected distribution of consumption for each country.

Moreover, a similar experience run on the French data but using the 2006 HBS wave proved to give better results, in particular in terms of fit with the expected distribution. The reason for such a result is that the distribution of food

consumption as measured in 2006 was far better closer to that observed in the HFCS data. As a result, the outcome of such an exercise remains pretty dependant on the good match in terms of distribution of food consumption between the two sources.

References

BLUNDELL, Richard; PISTAFERRI, Luigi; PRESTON, Ian. Imputing consumption in the PSID using food demand estimates from the CEX. 2004.

BROWNING, Martin; CROSSLEY, Thomas F.; WEBER, Guglielmo. Asking consumption questions in general purpose surveys*. The Economic Journal, 2003, 113. Jg., Nr. 491, S. F540-F567.

CIFALDI, Giulia; NERI, Andrea. Asking income and consumption questions in the same survey: what are the risks?. Bank of Italy Temi di Discussione (Working Paper) No, 2013, 908. Jg.

HAMERMESH, Daniel S. Consumption during retirement: The missing link in the life cycle. 1984.

HFCN: The Eurosystem Finance and Consumption Survey. Results from the First Wave. Statistics Paper Series, No. 1, 2013.

HFCN: The Eurosystem Finance and Consumption Survey. Methodological Report for the First Wave. Statistics Paper Series, No. 1, 2013.

SKINNER, Jonathan. A superior measure of consumption from the panel study of income dynamics. Economics Letters, 1987, 23. Jg., Nr. 2, S. 213-216.

Annexes

Annex 1: Questions on food consumption in the HFCS blueprint questionnaire

Let's now talk about household consumption:

9.01 HI0100

During last 12 months, about how much did (you/your household) spend in a typical month on food and beverages at home?

Numerical value in EUR, 6 digits. *-1 - Don't know*
-2 - No answer

9.02 HI0200

During the last 12 months, about how much did (you/your household) spend in a typical month on food and beverages outside the home? I mean expenses at restaurants, lunches, canteens, coffee shops and the like. Please, include only the amounts (you/your household) paid out i.e. net of any employer subsidy/discount/promotion etc.

Numerical value in EUR, 6 digits. *-1 - Don't know*
-2 - No answer

Annex 2: Comparison of the distribution of the covariates used in the model

Comparison of covariates

Spain — Table 4a

Covariates		HBS 2008	HFCS 2010	Gap
Food consumption	mean	7626.58	6416.04	-15.8%
	p5	1577.55	2064.00	30.8%
	p10	2213.09	2400.00	8.4%
	p25	3837.89	3600.00	-6.1%
	p50	6477.04	6000.00	-7.3%
	p75	10018.36	7740.00	-22.7%
	p90	14223.05	10800.00	-24.0%
	p95	17558.96	14400.00	-17.9%

Comparison of distribution of covariates in HBS and HFCS Spanish data

Comparison of covariates (ctd.)

Spain

Table 4a

Covariates		HBS 2008	HFCS 2010	Gap
Rents	mean	4687.17	4957.44	5.7%
	p5	342.24	360.00	5.1%
	p10	700.00	756.00	7.9%
	p25	2207.28	2520.00	14.1%
	p50	4386.84	4800.00	9.4%
	p75	6600.00	6876.00	4.1%
	p90	8760.00	9360.00	6.8%
	p95	10091.23	10800.00	7.0%
Age of RP	20-	0.1%	0.1%	-0.0%
	21-30	7.2%	6.0%	-1.2%
	31-40	21.3%	22.2%	0.9%
	41-50	21.3%	22.1%	0.7%
	51-60	17.1%	17.2%	0.1%
	61-70	14.0%	14.1%	0.0%
	70+	18.7%	18.1%	-0.6%
Size of household	1	22.2%	18.3%	-3.8%
	2	28.4%	29.5%	1.0%
	3	21.4%	25.3%	3.8%
	4	19.3%	21.3%	2.0%
	5	6.1%	4.0%	-2.0%
	6	1.5%	0.9%	-0.6%
	7	0.4%	0.1%	-0.2%
	8	0.1%	0.0%	-0.1%
	9	0.1%	0.2%	0.0%
Number of children	0	50.7%	45.7%	-4.9%
	1	23.5%	27.3%	3.7%
	2	20.3%	22.3%	1.9%
	3+	5.3%	4.6%	-0.7%
Schooling (ISCED code)	1	30.5%	34.3%	3.7%
	2	38.0%	19.7%	-18.3%
	3	12.5%	19.7%	7.1%
	5	18.8%	26.2%	7.4%
Tenure status	Free use	5.2%	5.5%	0.2%
	Home owner	80.2%	82.6%	2.4%
	Tenant	14.4%	11.6%	-2.7%

Comparison of distribution of covariates in HBS and HFCS Spanish data

Comparison of covariates

France — Table 4b

Covariates		HBS 2010	HFCS 2010	Gap
Food consumption	mean	4730.83	5858.02	23.8%
At home	p5	162.00	1200.00	40.7%
	p10	532.00	1800.00	38.3%
	p25	1853.00	2400.00	29.5%
	p50	3857.00	4320.00	12.0%
	p75	6650.00	6000.00	-9.7%
	p90	9848.00	9600.00	-2.5%
	p95	12193.00	12000.00	-1.5%
Food consumption	mean	1607.52	1232.99	-23.2%
outside	p5	0.00	0.00	.
	p10	0.00	0.00	.
	p25	0.00	0.00	.
	p50	674.00	480.00	-28.7%
	p75	2172.00	1200.00	-44.7%
	p90	4459.00	3000.00	-32.7%
	p95	6295.00	4392.00	-30.2%
Rents	mean	4520.29	6551.93	44.9%
	p5	108.00	2592.00	300.0%
	p10	810.00	3360.00	314.8%
	p25	2532.00	4380.00	72.9%
	p50	4200.00	5772.00	37.4%
	p75	6000.00	7200.00	20.0%
	p90	8088.00	9360.00	15.7%
	p95	9600.00	10752.00	12.0%
Age of RP	20-	1.2%	1.2%	-0.0%
	21-30	12.4%	11.8%	-0.5%
	31-40	16.5%	17.6%	1.1%
	41-50	19.3%	17.8%	-1.5%
	51-60	17.9%	17.6%	-0.2%
	61-70	14.4%	14.7%	0.2%
	70+	17.9%	18.9%	0.9%

Comparison of distribution of covariates in HBS and HFCS French data

Comparison of covariates (ctd.)

France

Table 4b

Covariates		HBS 2010	HFCS 2010	Gap
Size of household	1	34.6%	35.2%	0.5%
	2	33.1%	32.5%	-0.5%
	3	13.7%	13.7%	0.0%
	4	12.7%	12.0%	-0.7%
	5	4.3%	4.7%	0.4%
	6	0.9%	1.2%	0.2%
	7	0.2%	0.2%	0.0%
	8	0.0%	0.0%	-0.0%
	9	0.0%	0.0%	0.0%
Number of children	0	64.7%	64.6%	-0.0%
	1	16.1%	15.8%	-0.2%
	2	13.6%	13.1%	-0.5%
	3+	5.4%	6.3%	0.8%
Schooling (ISCED code)	1	28.3%	31.6%	3.2%
	2	28.6%	6.1%	-22.4%
	3	15.5%	38.5%	23.0%
	5	27.3%	23.5%	-3.8%
Tenure status	Home owner	59.8%	61.4%	1.5%
	Tenant	40.1%	38.5%	-1.5%

Comparison of distribution of covariates in HBS and HFCS French data

Comparison of covariates

Italy

Table 4c

Covariates		HBS 2009	HFCS 2010	Gap
Food consumption	mean	6506.92	5981.01	-8.0%
	p5	1696.44	2400.00	41.4%
	p10	2291.40	2400.00	4.7%
	p25	3568.20	3600.00	0.8%
	p50	5611.56	6000.00	6.9%
	p75	8405.52	7200.00	-14.3%
	p90	11865.24	9600.00	-19.0%
	p95	14305.32	12000.00	-16.1%

Comparison of distribution of covariates in HBS and HFCS Italian data

Comparison of covariates (ctd.)

Italy — Table 4c

Covariates		HBS 2009	HFCS 2010	Gap
Rents	mean	4469.02	4392.90	-1.7%
	p5	912.00	792.00	-13.1%
	p10	1536.00	1200.00	-21.8%
	p25	2844.00	2520.00	-11.3%
	p50	4200.00	4200.00	0.0%
	p75	5640.00	6000.00	6.3%
	p90	7200.00	7320.00	1.6%
	p95	8400.00	8400.00	0.0%
Age of RP	20-	0.3%	0.1%	-0.1%
	21-30	4.1%	3.9%	-0.2%
	31-40	15.4%	14.3%	-1.0%
	41-50	21.6%	23.7%	2.1%
	51-60	19.4%	18.2%	-1.2%
	61-70	16.2%	16.5%	0.3%
	70+	22.7%	22.9%	0.2%
Size of household	1	30.0%	24.9%	-5.0%
	2	27.0%	30.4%	3.3%
	3	20.0%	19.4%	-0.5%
	4	17.7%	18.7%	0.9%
	5	4.1%	4.8%	0.7%
	6	0.7%	1.4%	0.7%
	7	0.1%	0.0%	-0.0%
	8	0.0%	0.0%	0.0%
	9	0.0%	.	.
Number of children	0	53.6%	52.6%	-0.9%
	1	23.3%	22.2%	-1.1%
	2	18.7%	19.5%	0.8%
	3+	4.2%	5.4%	1.2%
Schooling (ISCED code)	1	26.8%	24.8%	-2.0%
	2	36.0%	28.4%	-7.5%
	3	25.8%	34.9%	9.0%
	5	11.2%	11.7%	0.5%
Tenure status	Free use	8.5%	10.2%	1.6%
	Home owner	74.2%	68.7%	-5.5%
	Tenant	17.1%	21.0%	3.9%

Comparison of distribution of covariates in HBS and HFCS Italian data

Annex 3: Estimation of the models on HBS data

Model for non-durable consumption

Spain

Table 5a

	Covariates	Coefficients
	Constant	7.10***
Food consumption	Log of the food consumption	0.25***
	Log2 of the food consumption	-0.04***
	Log3 of the food consumption	0.01***
Income deciles	2nd	1.63***
	3rd	11.44***
	4th	-0.39
	5th	2.40***
Rents	Log of rent	0.10***
	Log2 of rent	-0.03***
	Log3 of rent	0.00***
Age of the RP	RP less than 30	0.02*
	RP between 30 and 40	0.00
	RP between 40 and 50	Ref.
	RP between 50 and 60	0.01
	RP between 60 and 70	-0.01
	RP more than 70	-0.09***
Gender of the RP	RP male?	-0.04***
Household size	1 person in the HH	-0.08***
	2 person in the HH	Ref.
	1 person in the HH	0.01
Number of children	One child	0.04***
	2 children	0.08***
	3+ children	0.12***
Tenure status	Home Owner	0.06*
	Free use	0.02
	Tenant	Ref.
Diploma of the RP	ISCED 0+1	-0.12***
	ISCED 2	-0.04***
	ISCED 3+4	Ref.
	ISCED 5+6	0.08***
+Interactions between income deciles and food consumption		

Coefficients estimated thanks to OLS on HBS data

Model for non-durable consumption

France

Table 5b

	Covariates	Coefficients
	Constant	8.56***
Food consumption at home	Log of the food consumption	0.24***
	Log2 of the food consumption	-0.08***
	Log3 of the food consumption	0.01***
Income deciles	2nd	0.09**
	3rd	0.25***
	4th	0.42***
	5th	0.59***
Food consumption outside	Log of the food consumption outside	0.12***
	Log2 of the food consumption outside	-0.04***
	Log3 of the food consumption outside	0.00***
Utilities	Log of the utilities	0.19***
	Log2 of the utilities	-0.05***
	Log3 of the utilities	0.00***
Rents	Log of rent	0.11***
	Log2 of rent	-0.03***
	Log3 of rent	0.00***
Age of the RP	RP less than 30	0.03***
	RP between 30 and 40	0.01
	RP between 40 and 50	Ref.
	RP between 50 and 60	0.01
	RP between 60 and 70	0.03***
	RP more than 70	-0.01
Gender of the RP	RP male?	-0.01*
Household size	1 person in the HH	-0.06***
	2 person in the HH	Ref.
	1 person in the HH	0.02
Number of children	One child	0.02
	2 children	0.01
	3+ children	0.00
Tenure status	Home Owner	-0.12***
	Tenant	Ref.
Diploma of the RP	ISCED 0+1	-0.10***
	ISCED 2	-0.04***
	ISCED 3+4	Ref.
	ISCED 5+6	0.04***
+Interactions between income deciles and food consumption		

Coefficients estimated thanks to OLS on HBS data

Model for non-durable consumption

Italy

Table 5c

	Covariates	Coefficients
	Constant	3.74
Food consumption	Log of the food consumption	1.29***
	Log2 of the food consumption	-0.59***
	Log3 of the food consumption	0.08***
	Log4 of the food consumption	0.00***
Rents	Log of rent	2.46
	Log2 of rent	-0.32
	Log3 of rent	0.01
	Log4 of rent	0.00
Age of the RP	RP less than 30	-0.04*
	RP between 30 and 40	0.01
	RP between 40 and 50	Ref.
	RP between 50 and 60	0
	RP between 60 and 70	-0.05***
	RP more than 70	-0.16***
Gender of the RP	RP male?	0.00
Household size	1 person in the HH	-0.11***
	2 person in the HH	Ref.
	1 person in the HH	0.01
Number of children	One child	0.04***
	2 children	0.01
	3+ children	0.01
Tenure status	Home Owner	6.44
	Free use	6.33
	Tenant	Ref.
Diploma of the RP	ISCED 0+1	-0.24***
	ISCED 2	-0.10***
	ISCED 3+4	Ref.
	ISCED 5+6	0.13***
+Interactions between income deciles and food consumption		

Coefficients estimated thanks to OLS on HBS data