Income and consumption expenditure by households groups in National accounts*

Alessandra Coli
University of Pisa, Department of Statistics and Applied Mathematics

Francesca Tartamella
National Accounts Office - Istat

Abstract

In this paper we aim at estimating some significant items of the Italian National Accounts (NA henceforth) by households groups using micro survey data and NA figures simultaneously. Particularly we attempt to estimate disposable income and consumption expenditure by households groups, identified by social or economic characteristics of the family.

We take NA macro data as the top to be disaggregated by groups of households according to patterns coming from micro data sources on income and consumption expenditure. The crucial question is how to reconcile NA macro data with micro data from surveys on households budgets. As a matter of fact both approaches (macro and micro) try to measure the same phenomena but figures are often considerably different. Part of the paper is devoted to the comparison of NA and household budget surveys macro data in order to point out the main differences.

Surveys on households budgets do not always provide unambiguous information on overlapping or economic related monetary variables. As a consequence the disaggregation of NA aggregates moving from micro data sources may lead to incoherent results. We propose here to apply statistical matching in order to get a set of coherent data on households budgets. In particular we apply statistical matching to merge the Bank of Italy Survey on Household Income and Wealth (SHIW henceforth) and the ISTAT Survey on Households Budgets (HBS) for the years 2002 and 2004. For this last year we try to match the HBS and the European Survey on income and living conditions (EU-SILC) as well. The matched data set provide indicators for splitting NA income and consumption expenditure by household groups.

JEL codes: E01, C40

Keywords: micro-macro reconciliation, households consumption propensities, statistical matching.

* This paper was prepared for the 30th General Conference of The International Association for Research in Income and Wealth, Portoroz, Slovenia, 24 August to 30 August 2008

Introduction

National accounts statistics are the result of the integration of several data sources. Particularly the Italian National Accounts office uses (a variant of) the methodology by Stone, Champernowne and Meade to merge data coming from different data sources in a coherent framework. The method allows to move from an initial set of estimates based on independent methods and data sources to a final set of estimates, coherent (balanced) with the supply and use table accounting constraints. Unfortunately, for the time being, microeconomic statistics on households income are not included in the database underlying the balancing process, i.e. households income micro statistics *do not* enter the estimation process of the Italian NA.

The use of micro income statistics is essential for estimating income by Households sub-sectors. The best would be probably to introduce income micro statistics in the integration process underlying the building of NA. The supply and use table (i.e. the framework where integration takes place) would then be substituted by a Social accounting matrix with households income variables mainly estimated on the basis of micro income statistics.

For the time being we move from a context where NA income data do not match with income data collected by sample surveys on households budgets. In Italy several studies have been carried out to explain such discrepancies, mainly in the context of studies aimed at assessing the quality of income surveys (e.g. Brandolini, 1999). In the first part of the paper we take up this issue: substantive changes in Italian NA on one hand and the emergence of the European survey on income and living conditions (EU-SILC) on the other, have significantly modified the context. We take into account the Istat survey on Income (EU-SILC) and The Bank of Italy Survey on Income and Wealth (SHIW). We compare concepts and figures on income in order to point out the main differences with respect to NA values. For consumption expenditure we compare NA aggregates with aggregates derived from the Istat Survey on Households Budgets (HBS). This last data source is a fundamental input for the building of the Italian NA. Several studies have already been carried out on differences between HBS and NA (e.g. Istat 2000). Here we recall the main conclusions and update data comparisons.

The aim of this research is the estimate of disposable income and final consumption expenditure by groups of households within the boundaries of NA. We take NA macro data as the top to be disaggregated by groups of households according to patterns coming from micro data sources on income and consumption expenditure. SHIW and EU-SILC, once adjusted to NA definitions and concepts, provide patterns to distribute income by households group. On the other hand HBS provides the best information for distributing NA consumption expenditure by households groups. As it is well known surveys on income do not collect detailed information on consumption expenditure, while surveys on consumption usually include only few generic questions on income. The HBS by itself would not allow to group households according to income related characters like the household's main source of income. On the other hand, the exclusive use of SHIW (or EU-SILC) would preclude the analysis of consumption of households groups by categories of products or by function of consumption (which is an essential piece of information for the building of the SAM).

Households surveys do not always provide unambiguous information on overlapping monetary variables once these have been harmonized in definitions and classifications. As a consequence, inconsistencies may arise in the indicators to be used for distributing NA income and consumption expenditure by households group. This is particularly true when households are grouped by characters not directly collected from both data sources, especially if monetary. In National accounts each aggregate is coherent and consistent with the others. The estimate of disposable income and consumption expenditure by households' groups within NA boundaries must pursue such inner consistency. To this end it is necessary to merge income and consumption micro data before using such information for disaggregating NA macro aggregates.

A relevant issue is therefore how to match micro data coming from income surveys on one hand and consumption surveys on the other. We suggest to use statistical matching. Part II is devoted to the description of the method. We apply statistical matching to merge the Bank of Italy Survey on Household Income and Wealth (SHIW) and the ISTAT Survey on Households Budgets (HBS) for the years 2002 and 2004. For 2004 we present a preliminary statistical matching between the European survey on income and living conditions (EU-SILC) and HBS. The result of the matching is a "matched sample". This data set provides estimates of income and consumption expenditure from the micro perspective. The structure of income and consumption expenditure by households' groups is applied to NA aggregates. Results are shown in Part III.

PART 1 National accounts aggregates and surveys estimates

Comparisons between NA aggregates and surveys estimates are currently presented in official publications. More detailed analysis are presented in researches aimed at assessing the quality of sample surveys (e.g. Brandolini, 1999) or at validating input data for microsimulation models. In both cases National accounts aggregates represent an external yardstick.
Table 1 and table 2 present a rough comparison among macro aggregates. The comparison shows a sensitive difference between figures currently labelled with the same name. The SHIW and EU-SILC estimates of disposable income are significantly lower with respect to the NA counterpart aggregate (about 30% and 20% respectively). Households consumption expenditure from HBS is around 80% of the NA aggregate.

Table 1 – Households disposable income in SHIW, EU-SILC and National Accounts
Italy, 2004-mln current euro-

	SHIW	Eusilc	National accounts
disposable income per household	29,483	33,133	40,254
total disposable income	687,267	772,350	929,589
Shiw (Eusilc)/NA	73.93%	83.09%	100.00%

Table 2 - Households consumption expenditure in HBS and National Accounts
- Italy, 2004 - mln current euro –

	HBS	National accounts
consumption expenditure per household	28,572	35,081
total consumption expenditure	666,031	810,148
HBS/NA	82.21%	100.00%

1. Differences in households disposable income

Differences between NA households disposable income and disposable income estimated from income surveys(SHIW and EU-SILC) can be explained by three main reasons:

- The definition of Households is different in NA and in sample surveys (§ 1.1)
- Disposable income includes different income components in the compared data sources (§ 1.2)
- Surveys underestimate income because of the selection bias and the under-reporting of certain types of income (§ 1.3)

1.1 Households in National accounts and in the Surveys

National accounts examine the economic aggregates according to the subjects involved in the monetary and financial transactions, called *Institutional units*. They are elementary decision-making centres, characterised by uniformity of behaviour and decision-making autonomy in the exercise of their main function.

For analytical reasons, units are combined into groups (the so called Institutional sectors) on the basis of their main function. Three principal kinds of functions are distinguished: production, redistribution of income and consumption.

According to the present European System of National and Regional Accounts (ESA95 for short, see Eurostat, 1996), the Households sector behaves mainly as a consumption unit. Households as consumers are defined by the ESA95 as a "*small groups of persons who share the same living accommodations, who pool, some or all, of the income and wealth and who consume certain types of goods and services collectively, mainly housing and food"* (ESA95 § 2.75).

The sector is composed by the following units:

1. Individuals or groups of individuals whose main function is consumption and for own final use production;

2. Entrepreneurs whose economic behaviour cannot be separated from the economic behaviour of their households;

3. Persons living permanently in institutions for religious, health or other reasons (long-term patients in hospital, prisoners serving long sentences);

4. Non-profit institution (NPI) characterised by a non-relevant economic weight.

The first two categories cover a population approximately equal to the population analysed by the households surveys.

The weight of the second category is extremely relevant in the Italian national accounts, given the relevant number of small unincorporated enterprises in the Italian economy. As a consequence, the Household sector is a mix of consumer and productive units. Consumption is not the sector's main function. On the other side the Corporation sector neglects a considerable number of small enterprises. This mixture of productive and consumption functions within the Households sector has negative effects both for the analysis of enterprises' production and for the study of households' expenditure. For this reason the Italian official statistics provide separate accounts for *Consumer households* and the *Producer households*.

According to Istat assumptions, Producer households include:

- non financial sole proprietorships, informal and de facto partnerships with 5 or less employees;
- auxiliary financial sole proprietorships, informal and de facto partnerships with no employees.

Productive units which do not have the legal status of corporations, but are not classified among producer households either, are considered as quasi-corporations and are included in the corporations sector S11.

The SHIW, HBS and EU-SILC surveys select their samples from the registry office records. The survey unit is the so called legal family. A household is defined as a group of persons living permanently in the same house and pooling all or part of their incomes. This definition corresponds approximately to the groups 1 and 2 of the NA definition.

Survey data do not cover people living in institutions. Nevertheless, correcting for institutional households would have a negligible effect: according to 2001 Italian Census only the 0.90 per cent of the total resident population lives in institutions.

Surveys also disregard Non-profit institution (NPI) characterised by a non-relevant economic weight. According to the Italian national accounts, these units correspond only to non profit institutions organised informally. The effect of these units on the Households income is negligible, in that the disposable income of NPISH recognized as independent legal entities was around the 0,50% of Households disposable income in the years 2000 – 2006.

Finally surveys do not cover individuals who do not appear in registers or do not live permanently in a house (not regular, homeless).

In conclusion differences in the definition of Households explain only a small part of the discrepancy between NA aggregates and surveys estimates.

1.2 Differences in the definition of disposable income.

The next step is to compare definitions of monetary variables. For ESA95 disposable income is the balancing item of the secondary distribution of income account, it comes from summing operating surplus and mixed income, compensation of employees, property incomes (receivable minus payable), social benefits and current transfers and deducting current taxation, social contributions and payable current transfers.

Both SHIW and EU-SILC compute disposable income as sum of income components, asking questions about different types of income: labour income, capital income and transfers, *net of taxes and social contributions*.

The main difference in definitions with respect to NA is about self employed income: in surveys a unique item called “self employed income” is recorded for those declaring themselves as self employed (employer, own account worker, family worker) regardless of the legal status of the firm where they work. National accounts, instead, lack an aggregate defined as self employed income, since compensation of self employed is recorded in three different aggregates, according to legal status and dimension of the firm where they work:

1. Mixed income from producer households, i.e. compensation of self employed working in producer households
2. Withdrawals from quasi-corporations (D422 item in ESA95) , i.e. compensation of self employed working in corporations that “keep a complete set of accounts and

have no independent legal status. However, they have an economic financial behaviour that is different from that of their owners and similar to that of corporations" (ESA95 §2.13f)

3. Withdrawals from corporations (D423 item in ESA95) , i.e. compensation of self employed working in corporations.

For what concerns the SHIW, the above mentioned components have been estimated and compared with NA (see table 3a). Due to the lack of information the comparison between EU-SILC and SHIW is carried out only for the self employed category as a whole (see table 3b).

Various further differences between SHIW, EU-SILC and National Accounts involve the content and definitions of the income categories which add up to disposable income. The Appendix B describes the fine details of each of these items.

To compare NA and surveys income components it is necessary to estimate NA disposable income components net of taxes and social contributions. Not all incomes face the same tax rate and tax system; moreover, not all incomes face the same proportions of underground economy and therefore the same ex-post actual tax rates, even if the ex ante tax rate is the same. In a first attempt to compare SHIW and NA figures, we try to approach NA net values deducting taxation from NA incomes. We compute taxation on received interests (which has a proportional taxation rate of 12.5%), and we breakdown the remaining taxation proportionally on all incomes (even if the actual taxation system is progressive). A microsimulation model should be used to correctly allocate taxation among different income categories[1].

Table 3a and 3b show the results: two columns are displayed for NA values: column A reports gross values once concepts have been harmonized as explained in more detail in the appendix (for example wages and salaries are reported net of employees social contributions), column B reports "net" values: current taxation is deducted from wages and salaries, self employed incomes and property incomes.

Table 3a Households sector disposable income compared with SHIW figures– mln current euro.

	2004			
	Sector accounts		SHIW*	SHIW/NA(net)
	A (gross of taxation)	B (net of taxation)		
Wages and salaries net of social contributions paid by employees (D11-D6112)	370,005	318,457	267,962	84.1
Mixed income from producer households net of social contribution paid by self employed (part of S16 B2-D6113)	164,392	141,489	117,913	83.3
Gross operating surplus (B2)	92,578	79,680	126,393	158.6
Withdrawals from the income of quasi-corporatios (D422)	58,371	50,239	15,415	30.7
Social benefits (D62)	258,928	222,855	157,987	70.9
Dividends +other income from corporations (D421+D423)	78,157	67,268	16,994	25.3
Interests (recieved-paid) (D41) not corrected for Fisim	36,884	30,448	3,621	11.9
Current taxation on income (D5)	149,952			
Other proprerty income and transfers (net)	7,701	6,628	5,513	83.2
Disposable income (interests not corrected for Fisim)	**917,064**	**917,064**	**711,799**	**77.6**

*Net figures, sample weights have been used to expand to population.

[1] When the focus is on households sub sectors it is important to correctly allocate taxation (and social contributions) not only among income type, but also among households. For this purpose the use of a microsimulation model would be very useful.

Table 3b Households sector disposable income compared with EU-SILC figures– mln current euro.

	Sector accounts		Eusilc	Eusilc/NA(net)
	A (gross of taxation)	B (net of taxation)		
Wages and salaries net of social contributions paid by employees (D11-D6112)	370.005	318.457	308.192	96,8
Self employment income (Mixed income from producer households+D422+D423-D6113)	274.900	240.066	138.835	57,8
Gross operating surplus (B2)	92.578	79.680	131.286	164,8
Social benefits (D62)	258.928	222.855	199.378	89,5
Other capital incomes, net (Interests not corrected for Fisim)	76.917	65.408	10.818	16,5
Current taxation on income (D5)	149.952			
Other transfers (received)	24.380	21.291	5.305	24,9
Other transfers (paid)	- 30.692	- 30.692		
Disposable income (interests not corrected for Fisim)	**917.064**	**917.064**	**793.814**	**86,6**

Table 3a and table 3b show that both EU-SILC and SHIW underestimate households disposable income. In general the coverage is higher for wages and salaries and social security transfers, whereas it is quite low for self employed incomes and for property incomes.

Under coverage of wages and salaries, self employed income and property incomes is due partly to selection bias, partly to underreporting and non reporting phenomena. National accounts include an estimate of the underground economy which cannot be accounted for by income surveys: this can partly explain surveys under-coverage of wages and salaries and self employed incomes.

With respect to self employed income, part of the under coverage of surveys can be due to a different classification of actual and imputed rents. In NA actual rents are included in mixed income while imputed rents are part of the households operating surplus. Since surveys underestimate mixed income and overestimate operating surplus we may assume that the two aggregates counterbalance. Indeed, when answering to a survey households tend to state that secondary dwellings are not let, therefore imputed rent is computed instead of actual rents. In the case of SHIW , which separately records imputed rents on secondary dwellings, if the latter are considered as actual rent and included in mixed income, coverage of mixed income rises up to almost 90%, and over coverage of imputed rent considerably decreases to 106%. EU-SILC too overestimates imputed rents, especially if one considers that the reported values account only for owner-occupied dwellings. We can conclude that households seem to overvalue their property.

Underestimation of property income is similar in both surveys. This could partly be explained by selection bias and difficulties in sampling wealthy households, since financial wealth is generally more concentrated than labour income[2]. Several papers (Cannari and D’Alessio, 1993; Brandolini, 1999) document how SHIW estimates fall considerably short of corresponding flow of funds figures[3].

Not surprisingly, both surveys are not able to account for current transfers, since not all categories of this item are reported in surveys, as explained in detail in the appendix B, but

[2] Property income is computed multiplying declared stocks by appropriate interest rates.

[3] “Transaction and savings accounts appear to be underestimated in the SHIW by an average of 64 per cent, government bonds by 70 per cent, and private bonds, company shares and investment shares by 85 per cent; worryingly, the underestimation varies considerably from one year to the other” Brandolini A., Cannari L., D'Alessio G., Faiella I. (2004),

this is not a big concern, since this aggregates counts for less than 1% of total disposable income.

Table 3a and 3b show that SHIW can account for the 76% of disposable income while EU-SILC shows a better coverage, accounting for almost 87% of disposable income. The better result of EU-SILC in terms of total coverage is due to higher over-coverage of imputed rents and to better coverage of wages and salaries and social benefits Also EU-SILC seems to be unable to correctly represent property income and other transfers.

1.3 The impact of the selection bias and of the under-reporting

NA and surveys figures are not totally comparable, even when aggregates have been fully harmonized. We try here to subtract to NA values the effect of underground economy, therefore all differences left between NA and the survey should be accounted for selection bias and for underreporting of specific flows (for example interests).

Table 4 compares surveys amounts with NA figures net of taxation and unregistered economy. For SHIW figures we recomputed gross operating surplus counting imputed rents only for residence dwellings, while imputed rents on secondary dwellings have been summed to mixed income. EU-SILC instead computes imputed rents only for those dwellings used as a main residence by the households, therefore it is not possible to operate a similar correction.

Table 4 Households sector disposable income, net of taxation and underground economy: comparison with SHIW and EU-SILC figures.

	SHIW/NA net of taxes and undergound economy	Eusilc/NA net of taxes and undergound economy
Wages and salaries net of social contributions paid by employees (D11-D6112)	95.9	110.3
Self employed income	95.0	93.0
Gross operating surplus (B2)/imputed rents	106.7	132.8
Social benefits (D62)	74.5	91.7
Other property income (net)	17.1	16.5
Other transfers (net)	85.2	82.0
Disposable income (interests not corrected for Fisim)	88.3	101.3

Once NA figures have been reported net of the underground economy survey figures coverage considerably improves, at least for labour incomes. For SHIW, we could compute these amounts for each typology of self employment incomes and we found a coverage approaching 100% for mixed income, 80% for withdrawals from quasi-corporations, 66% for other income from corporations. It appears therefore that interviewed tend to have a coherent non-reporting or under-reporting behaviours when answering to households or to business surveys or filling social security registers, which are the bases of NA estimates.

2. Comparison between HBS estimates and NA consumption expenditure

HBS data are a relevant, albeit not unique, input for the estimate of consumption expenditure in the Italian National accounts[4].

In HBS households are asked to specify their expenditure on a quite detailed list of variables of all types of products. The list of products was decided on the basis of international classifications (Coicop) and agreed also with national accountants. The main conceptual difference between consumption from HBS and NA is that the first refers to residents expenditure while the NA aggregate refers to domestic consumption. Therefore in order to compare figures one should deduct from HBS consumption those expenditures made abroad by resident units and sum consumption made on national territory by non resident units.[5]

Furthermore some items considered in HBS total expenditure are not included in NA consumption:

- Expenditure in second hand durable goods. NA methodology is based on balancing consumption and investment on the one hand with production on the other:. Since exchange of second hand goods does not correspond to production in NA, second hand durable goods are not included in consumption or investment, while are in HBS expenditure.

- Non routine dwelling maintenance and repairs are excluded from NA consumption because they are considered as investment

- Life insurance premium, mortgage and other debt payback (capital and interests): these items are recorded among expenditures in HBS, while they are considered as saving by NA definitions.

- Holydays, hotels and restaurants paid abroad: they are recorded among HBS expenditures but not included in NA national consumption

Table 5 shows the percentage of NA consumption expenditure by functions covered by HBS data in 2004. Expenditure made abroad have been deducted from HBS figures, but NA domestic consumption contains expenditure of non resident units on the Italian territory

---

[4] Istat (2000).

[5] Consumption categories analysed in the survey can be grouped in:
– Food and beverages, tobacco (Coicop 01-02)
–Clothing and footwear (Coicop 03), in NA it includes as well dry cleaning and repairs
–Housing, water, electricity, gas and other fuels (Coicop 04). It includes all expenditure listed by HBS in this item for residence house and secondary houses (including imputed rents), except non ordinary maintenance and repairs.
–Furnishings, household equipment and routine household maintenance (Coicop 05)
–Health (Coicop 06)
–Transport and Communications (Coicop 07-08). It includes purchase of cars and transports. HBS asks to specify purchase of new or second hand products since only new transports are included in consumption.
–Recreation , culture and education (Coicop 09-10)
–Restaurants and hotels ((Coicop 11). HBS details expenditure made in Italy or abroad, since the latter is not included in domestic consumption.
Miscellaneous goods and services (Coicop 12). NA figure it includes Fisim, which is not computed in HBS. It includes as well consumption of insurance services, computed accounting paid premiums less claims received, while HBS reports only premium.

(28811 millions of euros), therefore figures are not fully comparable. That is why two columns have been reported for NA figures: column A showing Domestic expenditure, column B displaying something in between domestic and resident consumption: non resident units expenditure was deducted, but expenditures abroad of resident units was not summed up. In this way HBS and NA are fully comparable.

Note that, as for SHIW imputed rents, HBS housing expenditure is the only category where the survey figure is higher than in NA. This could confirm that households tend not to state that secondary dwellings are let, therefore imputed rent is computed instead of actual rent.

Table 5 Italian domestic expenditure for final consumption – mln current euro.

	2004				
	National Accounts			SHB/NA, %	
	A Domestic expenditure	B Domestic expenditure-expditure of non resident units	SHB	A over domestic expenditure	B over domestic expenditure-non resident unit
Food and beverages, tobacco	145,084	143,906	129,376	89.2	89.9
Clothing and footwear	67,825	64,291	46,033	67.9	71.6
Housing, water, electricity, gas and other fuels	166,209	166,209	184,377	110.9	110.9
Furnishings, household equipment and routine maintenance of the house	64,127	64,127	37,009	57.7	57.7
Health	26,613	26,613	24,704	92.8	92.8
Transport and Communications	134,413	129,111	84,226	62.7	65.2
Recreation , culture and education	66,883	63,643	40,254	60.2	63.2
Restaurants and hotels	80,597	65,039	28,520	35.4	43.9
Miscellaneous goods and services	74,944	74,944	49,946	66.6	66.6
TOTAL	**826,694**	**797,883**	**624,444**	**75.5**	**78.3**

3.Grossing up surveys figures to the NA aggregates

Households surveys underestimate income and consumption expenditure, therefore a method must be developed to gross up surveys values to the NA levels. The simplest solution is to revaluate each household total income and consumption by the same percentage by which surveys totals underestimate NA data. A more refined solution consists in applying specific revaluation coefficients for different income components and different kinds of consumption expenditure.

We have estimated specific revaluation coefficients for all income categories listed in table 3a, further detailing interests (applying therefore different coefficients to received and to paid interests) and mixed income (applying different coefficient to actual rents and to operating surplus from unincorporated enterprises[6]). The coefficient is computed as follows:

$$IC_I = \frac{1}{\frac{{}_S I_i}{{}_{NA} I_i}}$$

IC_i= revaluation coefficient for income component *i*

${}_S I_i$ = *i* income component value in surveys

${}_{NA} I_i$ = *i* income component value in NA

[6] Unpublished NA estimates.

Ideally one should apply coefficients that could be different for at least some of the socio-demographic-economic characteristics, but unfortunately we do not have any further information and we can just differentiate the coefficients by income categories. Availability of regional households accounts could help in computing grossing up coefficients that could at least incorporate one household characteristics, since it is likely that underreporting and non-reporting behaviour is not uniform by geographic area (as it is the case for underground economy). It would be advisable that Italian NA department start again publishing these estimates.

For what concerns consumption expenditure we have estimated revaluation coefficients for each consumption category.

$$CC_I = \frac{1}{\frac{{}_S C_i}{{}_{NA} C_i}}$$

CC_i= revaluation coefficient for the function of consumption *i*

${}_S I_i$ = consumption expenditure for the function of consumption *i* in HBS

${}_{NA} I_i$ = consumption expenditure for the function of consumption *i* in NA

PART II Matching microdata sets on households' income and consumption expenditure

As it is well known surveys on income do not collect detailed information on consumption expenditure, while surveys on consumption usually include only few generic questions on income. No single microdata set contains all of the different kinds of information that we are interested in. The exclusive use of the HBS would not allow to estimate income by groups of households. On the other hand, using the income surveys only would preclude the analysis of consumption of households groups by categories of products or by function of consumption (which is an essential piece of information for the building of the SAM).

However, the simultaneous use of independent data sources may lead to inconsistencies in the indicators to be used for distributing NA income and consumption expenditure by households group. This is particularly true when households are grouped by characters not directly collected by both data sources, especially if they are monetary.

It is therefore necessary to merge information collected by income surveys and consumption expenditure surveys. Several methods can be used to merge data drawn from independent data sources, such as calibration methods or linear regression models[7].

We suggest here to apply statistical matching in order to merge the SHIW and HBS microdata sets. The matching procedure is a complex and time consuming procedure since merging is performed at the single record (household) level. Our main focus is mainly on the coherence for groups of households (subsectors) rather than on the coherence of the matched file at the single household level. In this research we use matched data to estimate NA income and consumption expenditure by groups of Households. This is only the first step towards the building of a complete set of NA by Households subsectors. Nonetheless it is a very crucial step.

In this view statistical matching seems particular attractive with respect to other techniques such as linear regression modelling. Once income and consumption data sources have been matched, we can rely on a coherent and inner consistent set of information for splitting up most of NA variables (not only disposable income and total consumption expenditure) by households groups. Furthermore the matched file allows to group households according to a very large spectrum of characteristics.

In section 1 we describe the matching process applied to SHIW and HBS micro datasets (years 2002 and 2004). Section 2 is devoted to the description of statistical matching between EU-SILC and HBS data (year 2004).

From time to time, starting from the early seventies, many examples of microdata integration through statistical matching have been published (see Rässler 2002 for an exhaustive survey).

Statistical matching is used to link independent samples of data, A and B, by means of some variables common to both data files. The samples are extracted from the same population with relative sample sizes small enough as to ensure that the two samples have non empty intersection with negligible probability. Some variables Y appear only in A whereas some variables X appear only in B. A set of variables Z can be observed in both samples. Since usually a sample recording Y, Z and X at the same time does not exist, it is necessary to generate an artificial data set where each unit records Z, Y and X values.

[7]The Handbook on SAM (see Eurostat,2002) devotes one chapter on the issue of integration methods

The primary objective of matching is the study of the relationships between the variables not jointly observed X and Y.

Various methods can be used to match A and B. We refer here to the matching technique that uses the nearest neighbour matches.

According to this method statistical matching can be regarded as an imputation problem. Let us consider sample A as an incomplete data set where X variables are missing. A is then defined as the *recipient* sample. For every unit a_i , with i = (1,2,….,n_A), one x value from the observations of the donor sample B is selected. The donor unit b_j with j = (1,2,….n_B) is searched among the units belonging to B which present Z values ideally identical to those of the recipient unit a_i. The perfect match in terms of the common variables is not always possible. When this occurs the donor unit is selected on the basis of a distance measure d(Z). The donor unit is the unit with the smallest distance. When more donors are identified a random selection is performed.

The A and B files are merged in a single new and complete data set, $\tilde{A} = \{(x_1, z_1, \tilde{y}_1),......., (x_{nA}, z_{nA}, \tilde{y}_{nA})\}$. This artificial sample is considered representative of the true population of interest. Notice that $\tilde{A}$ has the same number of elements n_A as the recipient sample and that $\tilde{y}_i$ is the value of the donor unit belonging to B.

This method is relatively simple but it has a relevant undesirable implication. In effect, the application of traditional statistical matching implies the so called Conditional independence between Y and X given Z (see especially Rodgers 1984)). Conditional independence is produced for the variables not jointly observed even when such variables are conditionally dependent in reality. The problem is that often the relationship between the never jointly observed variables is unknown.

The Conditional independence Assumption (CIA) is a strong limit to the application of traditional statistical matching. Sceptics assert that statistical matching does not bring any additional information on the relationship between the not jointly observed variables: the outcome is already well known. The advocates argue that statistical matching is the only practical solution when the merging of data sets with hundreds of variables is necessary (see for example Ruggles 1974). According to this viewpoint, CIA can be roughly satisfied by carefully selecting the common variables (see Rässler 2002 for the debate on the pros and cons of statistical matching). For this reason it is essential to identify common variables significantly connected both with Y and Z variables.

1. the matching of SHIW and HBS data sets

The Italian National Institute survey on households budgets (HBS) focuses on households' expenditure. The survey also collects detailed information on the social, demographic and economic characteristics of the household (and/or of its components) as well as on the characteristics of the lodging where the family lives. Data on income and saving are collected in a specific section where households are asked to indicate the relevant interval for their monthly disposable income and annual saving. Such information is typically affected by underreporting. However, we will see how such variables are precious in the merging process. Unfortunately from the year 2003 onwards Istat does not release HBS data on income and saving

The Bank of Italy survey on Households Income and Wealth (SHIW) collects detailed information on income (especially on primary and capital income) and saving. It also contains questions on the demographic and socio-economic characters of the household's components and questions about lodging. Few generic question on Households expenditure for consumption are also included.

There is some consensus that SHIW supplies more reliable and detailed information on households' income while the HBS is more accurate in collecting data on households' expenditures. This depends obviously on the survey's specific objectives: the study of the distribution of income and wealth for SHIW and the analysis of households consumption's pattern for HBS.

The SHIW and HBS surveys cover approximately all the economic issues necessary to build Households' subsectors accounts. Unfortunately these surveys do not always provide unambiguous information on common monetary variables once these have been harmonized in definitions and classifications. For example the level and distribution of income may be significantly different, income from one source may not be economically coherent with consumption expenditure coming from the other source and vice-versa. The next table shows an example of the possible incoherent outcomes that can be obtained by using SHIW and HBS without any previous matching process.

Table 6 Consumption propensities by household regional area. Year 2004

	Consumption propensities (SHIW data)	Consumption propensities (HBS expenditure distribution on SHIW income distribution)
North-west	0.84	0.87
North-east	0.82	0.80
Centre	0.89	0,68
South	0.93	1.13

In the following we apply statistical matching to SHIW and HBS for the years 2002 and 2004[8].

Harmonizing SHIW and HBS common variables

As a first step, surveys have been harmonized in order to make the data comparable. This is a very time-consuming and difficult step. Inconsistencies must be solved through recoding of variables, imposing assumptions etc. It is necessary to carry out harmonization with care since changes on original data sets have relevant effects on the entire matching procedure (see Ruggles 1974)

Table 7 presents the list of the common variables with the indication of the harmonized categories.

[8] We replicate the method developed within an Istat working group. In the research studies for the building of a Social accounting matrix, the Italian national accounts office organised a working group aimed at integrating SHIW and HBS data sets. Full results are given in Coli et al. (2006).

Table 7 SHIW, HBS and EU-SILC common variables with harmonized categories

Name	Categories	Description	Name	Categories	Description
ndiploma	1=0 2=1 3=2 4=3 or more	Number of members with 11-13 years' schooling	**nanz**	1=0 2=1 3=2 or more	number of members aged >64
nlaurea	1=0 2=1 3=2 or more	Number of members with a university degree	**nadul**	1=0 2=1 3=2 or more	number of members aged 40-64
nobbligo	1=0 2=1 3=2 4=3 or more	Number of members with up to 8 years' schooling	**ngiova**	1=0 2=1 3=2 or more	number of members aged 18-39
qualim	1=1 2=2 3=3 4=4 5=5	quintile of expenditure for food consumption	**nmin**	1=0 2=1 3=2 4=3 or more	number of members aged < 18
nfem	1=0 2=1 3=2 4=3 or more	number of females	**tipoanz**	1=yes, 2=no	presence of at least one member aged >=75
ndip	1=0 2=1 3=2 or more	number of employees	**nocc**	1=0 2=1 3=2 or more	number of members with a job
nindip	1=0 2=1	number of self-employed	**mutup**	0=no 1=yes	the household is in charge of a house
naltrac	1=0 2=1 3=2 4=3 or more	number of members neither employed nor retired (housewifes, students etc)	**fitmut**	0=no 1=yes	the household pays a mortgage or a rent
npens	1=0 2=1 3=2 or more	number of retired members	**tbtr**	1=yes 2=no	house renting
nbam	1=0 2=1 3=2 4=3 or more	number of children aged < 15	**ncomp**	1,2,3,4,5 and more	number of members
tipobam	1=yes, 2=no	presence of at least one child aged < 15	**area**	1=North-Ovest 2=North-east 3=Centre 4=South	geogrphic area of residence

The head of the households characteristics have not been considered among the potential matching variables due to the different definitions of the head of the household (or reference person) in the three surveys. In fact the SHIW head of the household is the person who is in charge of the economic management of the family whereas the HBS and

EU-SILC reference person is the member indicated in the registry office records (on the basis of a pronouncement of the household) from which samples are extracted.

Table 8 Distribution of households by the head of the household's gender

SHIW, HBS and EU-SILC data - % values

	MALE	FEMALE
SHIW	60.97	39.03
SHIW(weighted)	61.19	38.81
HBS	73.31	26.69
HBS(weighted)	71.40	28.60
EU-SILC	71.70	28.30
EU-SILC(weighted)	71.50	28.50

Data in Table 8 seem to confirm such hypothesis. The EU-SILC and HBS surveys seem to identify the same member of the family as the head of the household, being the repartition between male and female approximately the same. On the contrary SHIW identifies a significant higher percentage of female heads of the household. Since definitions cannot be harmonized through the information available in the surveys, we decided not to consider these variables among the potential matching variables.

Sometimes it is not possible to use variables in the matching process even though their categories have been perfectly harmonized. This happens when the distributions of households by such variables are significantly different in the data sets, as if samples were extracted from different populations. Unfortunately this happens also for variables which it would be very desirable to include among the matching ones, given their strict link with income and consumption. In Appendix A, section 1 distributions are compared and a Chi square test is applied.

Different distributions are probably due to differences in sampling design, in the substitution of missing units, in the definitions of variables.

Contrary to our expectations, Chi square tests support the hypothesis that samples are extracted from populations where dimension and geographical residence of households are significantly *different*. The best results are obtained for education variables, followed by age variables, labour variables and housing variables.

Particularly, we notice the relevant differences in the distribution of households by number of retired members (***pens***) and by number of members neither employed nor retired (***naltrac***). Therefore we have decided to exclude these two variables from the set of the potential matching variables.

Selecting the matching variables

The second step consists in selecting the matching variables, i.e the common variables most strictly connected to household income and consumption. In order to satisfy the Conditional Independence Assumption as much as possible, it is essential to select variables that are strictly connected both with income and consumption expenditure. If we exclude monetary variables we have to resort to the demographic and social variables identified in the previous section.

It has already been pointed out that the HBS survey collects data on household income and saving. Particularly, it includes a categorical variable on the total monthly entries of the household. This variable underreports income but in our opinion it supplies a good piece of information on the rank of the households in the income cumulative distribution. For 2002 we have defined both for HBS and SHIW the new variable ***TM*** which classifies households according to their relative position in terms of income. Eight categories have been established, from the poorest household (TM=1) to the richest (TM=8). Each category contains approximately the same number of households. This variable gives a relevant contribution to the matching process. Unfortunately, as already pointed out, we cannot defined TM for 2004, since from 2003 onwards ISTAT does not supply any information on income and saving through the HBS.

As an alternative to the TM variable we have used the quintile of food consumption expenditure. Again we assume that both SHIW and HBS correctly collect the rank of the household in a range from lowest consumption to the the highest. .

On the basis of several analysis (see Appendix A, section 2) the following variables have been selected as the most strictly connected to households income and consumption.

TM: income class (only for 2002)
Qalim: quintile of food consumption expenditure
Ncomp: numbers of members
Nocc: numbers of members with a job
Ndip: number of employees
Ndiploma: number of members with 11-13 years' schooling
Nlaurea: Number of members with a university degree
Nadul: number of members aged 40-64
Tipoanz: presence of at least one member aged >=75
Area: geographical area of resident
Tbtr: house renting

Performing statistical matching

The SHIW microdata set is the recipient sample whereas the HBS microdata set is defined as the donor sample. The size of the SHIW sample is about 8000. Since the HBS sample is on average three times as large, more than one donor unit can be selected for each SHIW unit.

In the application of the nearest neighbour distance matching, different combinations of the matching variables can be chosen.

As a first step SHIW and HBS sample units are classified on the basis of the matching variables with the strongest relationship with income and consumption expenditure, the so called strata variables. The SHIW and HBS units are grouped into subsets (*strata*) whose units share the same exact value of the strata variables.

We have considered the following households stratifications of SHIW and HBS samples:

Year 2002

1) Income categories (***TM***) by Number of members (***Ncomp***): 40 strata
2) Quantile of consumption expenditure (***Qalim***) by Number of members (***Ncomp***): 25 strata

Year 2004

1) Quantile of consumption expenditure (***Qalim***) by Number of members (***Ncomp***): 25 strata
2) Number of members with a job (***Nocc***) by Numbers of members (***Ncomp***): 15 strata

The nearest neighbour of each SHIW unit is searched among the HBS units belonging to the same stratrum. As a first step the algorithm[9] identifies in each stratum the HBS units with the lowest distance from the SHIW recipient units where the distance is a function of the values assumed by the matching variables[10]. Whenever more than one donor is identified, a random selection is run.

Different combinations of the matching variables can be chosen in order to calculate distance. The following tables describe the performed combinations. The first column records the name of the data set which results from the matching process.

Tab 9 Combination of matching variables for SHIW-HBS statistical matching- 2002

	Maching variables	
Matched files	*Strata variables*	*Distance matching variables*
QNC	qalim,ncomp	nocc,ndiploma,nlaurea,ndip, nadul,tipoanz,tabt,area
TMNC	TM,ncomp	nocc,ndiploma,nlaurea,ndip, nadul,tipoanz,tabt,area

* see tab 7 for the definitions of variables

Tab 10 The combination of matching variables for SHIW-HBS statistical matching -2004

Matched files	**Maching variables**	
	Strata variables	*Distance matching variables*
QNC1	qalim,ncomp	nocc,ndiploma,nlaurea,ndip,nadul, tipoanz,tabt,area
QNC2	qalim,ncomp	nocc,ndiploma,nlaurea,ndip,nadul
QNC3	qalim,ncomp	nocc,nadul,ndiploma, ndip
QNO1	qalim,nocc	ncomp,ndiploma,nlaurea,ndip,nadul, tipoanz,tabt,area
QNO2	qalim,nocc	ncomp,ndiploma,nlaurea,ndip,nadul
QNO3	qalim,nocc	ncomp,nadul,ndiploma, ndip

* see tab 7 for the definitions of variables

The statistical matching procedure generates matched files which have the same dimension as the SHIW (recipient dataset). In order to choose the best matched file it is

[9] We use a software developed by Giuseppe Sacco (Istat); see Coli et al. 2006 for details.
[10] We have used Gower distance (see Gower,1971)

customary to compare the distributions of imputed variable in the donor and in the matched data set (see Rässler 2002, D'Orazio et al. 2006). As our main objective is to impute HBS consumption expenditure, we compare summary statistics on total consumption expenditure calculated with HBS data and with each matched file data.

Table 11 Comparisons between consumption expenditure statistics computed on each matched file data and on the HBS data (HBS statistic=100) – year 2002

Summary statistics	Matched data sets			
	QNC	TMNC	QNC	TMNC
	Unweighted values		Weighted values	
μ	99.02	98.34	99.55	100.20
σ	101.94	96.53	104.73	101.30

Tab 12 Comparisons between consumption expenditure statistics computed on each matched file data and on HBS data (HBS statistic=100) – year 2004

	Matched data sets					
Unweighted values						
	QNC1	QNC2	QNC3	QNO1	QNO2	QNO3
μ	102.55	102.72	104.06	103.18	104.39	103.66
σ	91.52	94.19	93.81	92.85	96.44	92.54
Weighted values						
	QNC1	QNC2	QNC3	QNO1	QNO2	QNO3
μ	106.95	105.79	108.58	108.57	106.83	107.30
σ	101.88	103.31	103.26	108.05	106.17	99.68

The summary characteristics of total consumption expenditure are quite well preserved in the 2002 matched files. The best result is obtained by the TMNC file (weighted values). In this file households are stratified by income category (***TM***) and by number of members variable (***ncomp***).

For what concerns the 2004 matched files, we notice a general tendency to underestimate the variability of consumption expenditure when unweighted values are considered. The contrary happens with weighted values. The mean is preserved quite well especially by the matched files obtained with households classified by quintile of consumption (***qalim***) and by number of members (***ncomp***).

Estimated correlations between imputed consumption ($\tilde{C}$) and observed SHIW income (Y) show clearly the effects of the Conditional independence assumption. Values are quite low if compared to the SHIW inner correlation between income and consumption which is around 0.60 in SHIW. The highest correlation is recorded for the TMNC file, followed by QNC, QNC1 e QNO1.

The QNC and TMNC matched data sets are obtained using the same combination of matching variables in the distance function. Strata variables are different: in QNC a stratum identifies households of the same dimension which belong to the same quintile of food consumption; in TMCN a stratum is composed by households of the same dimension belonging to the same income category. Substituting the income category (***TM***) for the quintile of food consumption (**qalim**) clearly leads to an increase of the correlation between imputed consumption and SHIW income.

For the year 2004, we notice how estimated correlations decrease when some matching variables are excluded from the computation of the distance.

Table 13 Correlations between imputed consumption ($\tilde{C}$) and SHIW income (Y) - 2002

	Matched data sets	
	QNC	**TMNC**
$_Y\rho_{\tilde{C}}$	0.329	0.390

Table 14 Correlations between imputed consumption ($\tilde{C}$) and SHIW income (Y)

	QNC1	**QNC2**	**QNC3**	**QNO1**	**QNO2**	**QNO3**
$_Y\rho_{\tilde{C}}$	0.308	0.273	0.255	0.310	0.296	0.267

On the basis of these results we chose the QNC1 file as the best one.
This file is the result of the statistical matching between SHIW and HBS data sets based on the stratification of households by quintile of food consumption and by number of members. This means that the HBS donor is a household of the same dimension as the recipient one, belonging to the same quintile of food consumption.

2. The merging of EU-SILC and HBS data sets

In this section we apply statistical matching to EU-SILC and HBS data sets, for the year 2004.

Matching variables can be selected only among the social and demographic characteristics of the household, since there are not comparable monetary variables in the surveys. Differently from SHIW, EU-SILC does not provide any information on households total or food consumption expenditure and HBS does not supply any information on income.

We made two trials: in the first households are stratified by geographic area (***Area***) and number of members (***Ncomp***), in the second, households are stratified by number of members (***Ncomp***) and number of members with a job (***Nocc***).

The two trials give similar results both in terms of preservation of the HBS consumption summary statistics and in terms of correlation between imputed consumption and EU-SILC income. Results are undoubtedly worse if compared to the SHIW-HBS matching.

Table 15 The combination of matching variables for EU-SILC-HBS statistical matching - 2004

	Maching variables	
Matched files	*Strata variables*	*Distance matching variables*
ARNC	area,ncomp	nocc,ndiploma,nlaurea,ndip, tipoanz,tabt,nadul
NONC	ncomp, nocc	area,ndiploma,nlaurea,ndip, tipoanz,tabt,nadul

Tab 16 Comparisons between consumption expenditure statistics computed on each matched file data and on HBS data (HBS statistic=100) – year 2004

Summary statistics	Matched data sets			
	ARNC	NONC	ARNC	NONC
	Unweighted values		Weighted values	
μ	113.86	114.56	115.32	115.55
σ	101.60	105.61	107.14	109.01

Table 17 Correlations between imputed consumption ($\tilde{C}$) and EU-SILC household income (Y) – 2004 –

	Matched data sets	
	ARNC	NOCN
$_{Y}\rho_{\tilde{C}}$	0.238	0.229

In order to fulfil CIA assumption as much as possible, it is essential to consider matching variables capable of explaining most of income and consumption variability. Given the available data, we have to rely on rather weakly explaining variables. The possibility of including variables such as the quintile of food consumption or the income categories considered for the SHIW – HBS matching would increase the quality of the matched file.

PART III Estimate of NA disposable income and consumption expenditure by Households' subsectors

In this last part of the paper we provide estimates of 2004 NA disposable income and consumption expenditure by households subgroups.

We subdivide NA variables according to indicators derived in the SHIW-HBS matched file (QNC1 file). In a first attempt we disaggregate NA disposable income according to the distributions by households subgroups of income and consumption as they result after the matching process. Income corresponds to the original SHIW value (Y) while consumption ($\tilde{C}$) is the HBS value imputed with the matching process. In a second attempt we use income and consumption expenditure distributions once QNC1 values have been grossed up to the NA values on the basis of the revaluation coefficients described in Part I (see Section 3).

In the second section an analysis of households consumption expenditures by households subgroups is provided.

1. Income and consumption expenditure by Households' subsectors

In Part II we have run several matching procedures for the year 2004, based on different households stratifications and different combinations of the matching variables for the computation of the function distance. Among the matched files we have identified file QCN1 as one of the best.

As a first step, we break down NA disposable income and consumption expenditure by households groups on the basis of indicators derived in QCN1. Indicators are SHIW disposable income (Y) and HBS imputed consumption ($\tilde{C}$). This method implies that differences in the level of surveys income and consumption with respect to the NA totals are distributed among households groups in proportion to their weights in the Y and $\tilde{C}$ distributions. Results are shown in table 18.

In a second attempt we have grossed up QNC1 income and consumption values on the basis of the revaluation coefficients described in Part I. For each household of the matched file, the various components of disposable income have been revaluated according to proper coefficients. Consumption expenditure as well has been revaluated with coefficients estimated at the functions of consumption level. We name as QNC1-R the matched file with grossed up values of income and consumption. NA aggregates are then broken down on the basis of QNC1-R figures. This method implies the use of more refined indicators: income and consumption expenditure are revaluated taking into account the income structure as well as the consumption pattern of each household.
In order to validate our estimates we have compared consumption propensities, computed on different kinds of income and consumption. Results are given in table 19 where four kinds of consumption propensities (CP) are compared. Columns names indicate the data source which provides the indicators for the computation of propensities. Propensities are computed as follows:

$CP_{(QNC1)}$=imputed consumption / SHIW income
$CP_{(\,QNC1\text{-}R)}$= imputed revaluated consumption /SHIW revaluated income
$CP_{(SHIW)}$ =SHIW consumption/SHIW income

$CP_{(HBS)}$ =HBS consumption/SHIW income.

Table 18 National accounts Disposable income and Consumption expenditure by households groups –Italy – year 2004 – mln current euro.

Households classified by geographical area **(Area)**

	North-East	**North-West**	**Centre**	**South**	**Total**
Disposable income	249,730	227,642	220,520	231,697	929,589
Consumption expenditure	222,374	196,285	177,088	214,401	810,148

Households classified by number of children **(Nbam)**

	1	**2**	**3**	**4+**	**Total**
Disposable income	714,249	125,920	76,897	12,523	929,589
expenditure	596,191	124,460	76,270	13,227	810,148

Households classified by number of retired people **(Npens)**

	0	**1**	**2**	**Total**
Disposable income	497,602	280,899	151,088	929,589
Consumption expenditure	457,293	232,401	120,454	810,148

Households classified by number of members with a job **(Ndip)**

	0	**1**	**2+**	**Total**
Disposable income	380,717	306,848	242,025	929,589
Consumption expenditure	308,443	278,936	222,769	810,148

Households classified by number of self-employed **(Nindip)**

	0	**1**	**2+**	**Total**
Disposable income	754,527	140,580	34,482	929,589
Consumption expenditure	672,852	114,490	22,806	810,148

Households classified by number of female members **(Nfem)**

	0	**1**	**2**	**3+**	**Total**
Disposable income	66,324	510,729	264,437	88,098	929,589
Consumption expenditure	45,983	437,969	241,125	85,071	810,148

Households classified by number of people with a University degree **(Nlaurea)**

	0	**1**	**2+**	**Total**
Disposable income	707,770	148,184	73,635	929,589
Consumption expenditure	639,384	120,695	50,069	810,148

Households classified by number of elderly people **(Nanz)**

	0	**1**	**2+**	**Total**
Disposable income	616,412	184,149	129,028	929,589
Consumption expenditure	563,896	141,055	105,197	810,148

Table 19 Consumption propensities for households groups. Propensities are computed on different values of household income and consumption expenditure.

	QNC1	QNC1-R	SHIW	SHB-SHIW
area				
1	89.0	87.6	84.4	86.9
2	86.2	83.8	82.3	79.9
3	80.3	80.8	89.0	68.2
4	92.5	95.7	93.2	112.6
nbam				
0	83.5	82.3	86.1	82.8
1	98.8	99.6	91.7	107.7
2	99.2	94.6	86.9	94.0
3	105.6	98.9	100.4	84.5
npens				
0	91.9	90.6	88.6	104.5
1	82.7	83	87.9	70.0
2	79.7	78.9	80.9	61.9
ndip				
0	81.0	73.7	89.3	76.9
1	90.9	91.8	87.2	97.2
2	92.0	103.1	83.8	90.5
nindip				
0	89.2	99.5	89.3	80.2
1	81.4	65.6	79.2	112.4
2	66.1	46.9	73.3	136.1
nfemm				
0	109.6	69.9	86.2	94.9
1	88.6	85.8	86.0	93.5
2	83.3	93.7	88.7	79.2
3	78.7	88.9	89.9	75.6
nlaurea				
0	90.3	93	89.8	90.3
1	81.4	68.5	80.1	79.2
2	68.0	59.3	75.5	72.7
nanz				
0	91.5	91.1	87.6	93.5
1	76.6	78.5	87.7	75.8
2	81.5	76.7	84.1	72.9

For most of the considered households subgroups QNC1 and QNC1-R propensities take more realistic values with respect to propensities calculated by using SHIW and HBS data without any previous matching process. Besides, propensities computed on the basis of the matched file values are more in line with those calculated with SHIW values.

The use of QNC1-R instead of QNC1 indicators has a significant impact on the subdivision of income and consumption by number of employed people, by number of self-employed people, by number of female members. Further investigation is required in order to explain differences.

2. Consumption expenditure by Households subgroups and consumption categories

Thanks to the detailed information in the matched file it is finally possible to analyse consumption expenditures by households subgroups. The analysis are based on the matched file QNC1-R with households income and consumption grossed up to NA totals through proper coefficients.

Comparing table 20 with table 5, Part I , it is worth stressing the higher coverage of the matched file values (before data are grossed up to NA values) for what concerns domestic expenditures of resident units for each consumption function.

Table 20 Italian domestic expenditure for final consumption – mln current euro.

	2004				
	National Accounts			SHIW-SHB/NA, %	
	A Domestic expenditure	B Domestic expenditure-expditure of non resident units	matched SHIW-SHB	A over domestic expenditure	B over domestic expenditure-non resident unit expenditure
Food and beverages, tobacco	145,084	143,906	136,886	94.3	95.1
Clothing and footwear	67,825	64,291	46,807	69.0	72.8
Housing, water, electricity, gas and other fuels	166,209	166,209	177,217	106.6	106.6
Furnishings, household equipment and routine maintenance of the house	64,127	64,127	42,817	66.8	66.8
Health	26,613	26,613	28,759	108.1	108.1
Transport and Communications	134,413	129,111	87,541	65.1	67.8
Recreation , culture and education	66,883	63,643	41,410	61.9	65.1
Restaurants and hotels	80,597	65,039	29,034	36.0	44.6
Miscellaneous goods and services	74,944	74,944	50,301	67.1	67.1
TOTAL	**826,694**	**797,883**	**640,771**	**77.5**	**80.3**

Table 21 displays consumption propensities for each expenditure function for the households subgroups analysed so far.
It is interesting to note that demographic variables have an impact on expenditures: while consumption propensities are increasing with respect to number of children aged less than 15 (***Nbam***), the opposite happens for the number of persons aged more than 65 years (***Nanz***)[11]. For both household groups expenditure in food and beverages grows with the number of children/old persons. Probably for households with children this is due to a budget constraint, whereas for households with old persons it is due to different consumption behaviour and preferences. Note in fact that “old households” tend to spend less also for clothing and house furniture, while “young households” spend more on these items as the number of children grows.

[11] The household typology characterised by number of recipients of social security incomes show a similar behaviour.

It is also worth remarking the different behaviour of households classified by number of employed members (*Ndip*) and by number of self employed members (Nindip). While consumption propensities tend to decrease with the number of self employed (as the average income increases), propensities increase with number of employees. As a matter of fact, while the incidence of each consumption function over disposable income increases with the number of employees, the opposite is true when the number of self employed increases.

Households by number of persons with a degree (***Nlaurea***) show a pattern similar to households with self employed, both in terms of consumption propensities and consumption pattern.

Expenditure composition is also different by geographic area: households living in the south have higher propensities to consume and the higher incidence of food consumption is striking. They also spend more for clothing and house equipment, less in restaurants and hotels.

Finally observe that the number of women (***Nfemm***) has also a significant impact. Households with no females (mainly households with only one component) have a much lower consumption in all expenditure functions, except housing. When analysing the incidence of each expenditure function over total consumption it appears that this kind of households have a much higher expenditure in restaurants and hotels.

Table 21: Incidence of consumption functions on disposable income by households subgroup

		Food and beverages, tobacco	Clothing and footwear	Housing, water, electricity, etc.	Furnishings household equipment etc.	Health	Transport and Communic.	Recreation, culture and education	Restaurants and hotels	Miscell. goods and services	*Consumption propensities*
nbam	0	14.9	6.4	18.3	6.1	2.7	13.3	6.3	6.5	7.8	*82.3*
	1	16.8	8.5	18.2	9.5	3.2	16.9	8.4	8.7	9.2	*99.6*
	2	17.4	8.4	17.6	8.1	3.3	14.4	8.3	8.3	8.9	*94.6*
	3	22.2	9.0	16.0	9.4	2.9	15.9	7.6	6.2	9.6	*98.9*
area	1	14.4	6.4	19.2	7.3	3.2	13.8	7.3	7.4	8.7	*87.6*
	2	12.6	6.5	17.2	6.8	2.9	15.2	6.8	8.0	7.8	*83.8*
	3	15.1	6.3	18.1	6.1	2.2	12.2	6.5	6.9	7.5	*80.8*
	4	21.3	9.1	17.7	7.7	3.3	15.1	7.1	5.9	8.6	*95.7*
ndip	0	14.8	5.1	18.5	5.0	2.9	10.6	5.2	4.9	6.8	*73.7*
	1	16.0	7.7	17.6	7.0	3.1	15.8	7.6	8.0	9.0	*91.8*
	2	16.8	9.3	18.2	10.4	2.6	17.6	9.0	9.7	9.4	*103.1*
npens	0	15.1	8.0	17.4	7.9	2.6	15.0	7.7	8.1	8.9	*90.6*
	1	16.5	5.7	20.0	5.9	3.2	13.0	5.8	5.6	7.4	*83.0*
	2	16.8	5.3	17.8	5.2	3.4	12.3	6.1	5.4	6.7	*78.9*
nfemm	0	10.9	5.0	18.8	4.1	2.0	11.0	5.6	6.3	6.3	*69.9*
	1	15.7	6.5	18.8	7.4	2.7	13.8	6.2	6.8	8.0	*85.8*
	2	16.5	8.3	17.5	6.9	3.4	15.8	8.4	7.9	9.0	*93.7*
	3	17.3	8.2	15.7	7.9	3.2	13.3	8.0	6.9	8.4	*88.9*
nindip	0	18.7	7.8	21.5	7.8	3.5	15.7	7.8	7.7	9.1	*99.5*
	1	10.2	5.7	12.3	6.1	1.7	11.3	5.4	6.3	6.7	*65.6*
	2	7.7	4.3	7.8	2.5	1.3	8.9	4.8	4.5	5.1	*46.9*
nanzi	0	15.4	7.9	17.5	7.7	2.8	15.3	7.6	8.1	8.8	*91.1*
	1	16.2	5.4	20.4	5.3	3.0	10.7	5.8	4.6	7.0	*78.5*
	2	16.5	4.5	18.3	5.5	3.4	12.3	4.9	4.8	6.5	*76.7*
nlaurea	0	17.5	7.4	19.3	7.3	3.2	15.3	7.1	7.3	8.6	*93.0*
	1	9.9	6.0	14.5	5.3	1.9	10.6	6.7	6.9	6.7	*68.5*
	2	8.0	4.9	12.6	7.9	1.2	8.1	5.6	4.8	6.1	*59.3*
National average		**15.7**	**7.0**	**18.1**	**7.0**	**2.9**	**14.1**	**6.9**	**7.1**	**8.2**	***87.0***

Concluding remarks

The ultimate purpose of this research is the estimate of disposable income and consumption expenditure by households groups within the National accounts framework. Estimates are shown in Part III of the paper where an analysis of consumption propensities and consumption patterns by households group is also provided.

The estimates of NA disposable income and consumption by households groups are the last step of a complex procedure. The crucial question is how to merge different data sources.

In the first part of the paper we concentrate on the comparison between NA on one side and households surveys on the other. The Istat survey on households consumptions expenditure (HBS) is fully integrated in the NA estimates. Income surveys (SHIW: the Bank of Italy survey on income and wealth and EU-SILC: the Istat survey on income and living conditions) on the contrary are not considered part of the Italian NA input data sources. Part I of the paper is devoted to the comparison of income and consumption expenditure in NA and in households surveys. The harmonization of definitions is not sufficient to eliminate differences in levels: households surveys fall short of NA income and consumption for a significant percentage. Comparisons at the level of the single income components show that self employed income, property income and transfer income record the lowest levels with respect to NA figures. For what concerns self employed income and wages and salaries, a significant percentage of the difference can be explained by the underground economy. In fact, once NA figures have been reported net of the underground economy, survey figures coverage considerably improves, approaching 100% for mixed income, 80% for withdrawals from quasi-corporations, 66% for other income from corporations. Comparisons of NA and HBS expenditures at the level of functions of consumption show that survey data underestimate expenditure for all categories of consumption, with levels of underestimation that vary significantly from one function to the other. Part I ends with the estimate of coefficients (revaluation coefficients) which must be applied to the survey data in order to reach the NA income and consumption expenditure levels. Revaluation coefficients have been estimated by income categories and by functions of consumption. The grossing up to NA values therefore takes into account variability in households income and consumption structures.

The second part of the paper is devoted to the comparison and fusion of households surveys data. The integration of survey data is at least as important as the integration of survey data in NA. In order to subdivide NA values by households subgroups it is essential to build a data set with complete and coherent information on monetary variables as well as on the social economic and demographic characteristics of the household. We have applied statistical matching in order to merge income surveys (SHIW and EU-SILC) with the HBS surveys. We preferred statistical matching over other merging techniques principally because our purpose is the linkage of a huge amount of variables. Several matching procedures have been run and the present paper describes only the most significant results.

The revaluation coefficients (Part I) and the matched file (Part II) are used to estimate NA income and consumption expenditure by households groups (Part III). The analysis of consumption propensities by households groups shows that the matched file supplies better indicators (for the disaggregation of NA values) than the not matched SHIW and HBS data sets. Further investigation is needed to explain differences in consumption propensities computed on the basis of the matched file data, or using the matched file variables once income and consumption expenditure have been grossed up to NA levels applying the revaluation coefficients.

References

Brandolini, A., 1999. *The Distribution of Personal Income in Post-War Italy: Source Description, Date Quality, and the Time Pattern of Income Inequality*, Banca Italia - Servizio di Studi.

Brandolini A., Cannari L., D'Alessio G., Faiella I. (2004), *Household Wealth Distribution in Italy in the 1990s*, Banca d'Italia, Temi di discussione n. 530.

Brandolini A., Cannari L., D'Alessio G., Faiella I. (2004), *Household Wealth Distribution in Italy in the 1990s*, Banca d'Italia, Temi di discussione n. 530.

Cannari I. D’alessio g. (1993) *Non-reporting and under-reporting behaviour in the Bank of Italy’s survey of household income and wealth* proceedings of the ISI 49th session _ Firenze 1993.

Coli A., S. Colombini, M. Di Zio, M. D’Orazio, I. Faiella, I. Siciliani, G. Sacco, M. Scanu, F. Tartamella (2006), *La costruzione di un Archivio di microdati sulle famiglie italiane ottenuto integrando l’indagine ISTAT sui consumi delle famiglie italiane e l’Indagine Banca d’Italia sui bilanci delle famiglie italiane*, Istat, Contributi.

D’Orazio M., M. Di Zio, M. Scanu (2005) *Statistical Matching and the Likelihood principle: Uncertainty and logical Constraints*, Istat, Contributi.

D’Orazio M., M. Di Zio, M. Scanu (2006), *Statistical matching: Theory and Practice*, Wiley & Sons, New York.

European Commission, International Monetary Fund, Organization for Economic Co-operation and Development, United Nations, World Bank (1994), *System of National Accounts* (SNA93), New York.

Eurostat (1996) European System of National and Regional Accounts (ESA 95), Luxembourg.

Eurostat (2002), Handbook on Social accounting matrices and labour accounts, Luxembourg.

Filippello R., U. Guarnera, Jona Lasino G., *Use of Auxiliary Information in Statistical Matching*, SIS, Riunione Scientifica - Bari 2004 – Comunicazioni spontanee.

Gower, J. C. (1971) A general coefficient of similarity and some of its properties. *Biometrika*

Istat (2000) *Le nuove stime dei consumi finali delle famiglie secondo il SEC 95*, Metodi e norme n. 7)

Istat (2005) *I conti economici nazionali per settore istituzionale:le nuove stime secondo il Sec95*, Metodi e norme n.23

Istat (2007) *Conti economici nazionali per settore istituzionale*, Statistiche in breve

Istat (2008) La misura dell'economia sommersa nelle statistiche ufficiali, Statistiche in breve

Kadane Joseph B. (2001) *Some statistical problems in merging data files* – Journal of official statistics, vol. 17, No. 3. Reprint (1978).

Moriarity C., Sheuren Fritz (2001) *Statistical matching: a paradigm for assessing the uncertainty in the procedure* – Journal of official statistics, vol. 17, No. 3.

Prem K. Goel, T. Ramalingam (1989), *The matching methodology: some statistical properties*, Lecture notes in statistics, 52 –Springer, New York.

Rodger W. L. (1984) *An evaluation of statistical matching* – Journal of business and economic statistics, vol.2, no. 1.

Rosati Nicoletta (1998) *Matching statistico tra I dati Istat sui consumi e dati bankitalia sui redditi per il 1995*. Rapporto tecnico.

Richard Ruggles, Nancy D. Ruggles (1974) *A strategy for merging and matching microdata sets*, Annals of economic and social measurement, 3, 1974.

Susanne Rässler (2002), *Statistical matching: a frequentist theory, practical applications and alternative Bayesian approaches*, Lecture notes in statistics, 168 – New York Springer.

Sutherland Holly, Rebecca Taylor and Joanna Gomulka (2001) *Combining household income and expenditure data in policy simulations*, Working paper Department of Applied economics, University of Cambridge.

Appendix A: Statistical analysis

A.1 Comparing samples

The matching between SHIW (or EU-SILC) and HBS microdata sets is based on common variables. Table A1 presents the empirical distributions of a set of common variables, once the categories have been harmonized.

Table A.1 Empirical distribution of households in SHIW, HBS and EU-SILC - 2004 -

Variable=ndiploma

	0	1	2	3+
SHIW	0.523	0.288	0.146	0.044
HBS	0.503	0.290	0.160	0.048
EU-SILC	0.496	0.311	0.152	0.041
SHIW(weighted)	0.514	0.294	0.147	0.044
HBS(weighted)	0.516	0.288	0.154	0.042
EU-SILC (weighted)	0.519	0.313	0.135	0.035

Variable=nlaurea

	0	1	2+
SHIW	0.857	0.106	0.037
HBS	0.854	0.108	0.039
EU-SILC	0.848	0.119	0.033
SHIW(weighted)	0.862	0. 104	0.034
HBS(weighted)	0.863	0.103	0.034
EU-SILC (weighted)	0.837	0.124	0.038

Variable=nobbligo

	0	1	2	3+
SHIW	0.17	0.31	0.32	0.20
HBS	0.17	0.31	0.31	0.21
EU-SILC	0.22	0.27	0.33	0.18
SHIW(weighted)	0.17	0.31	0.30	0.21
HBS(weighted)	0.18	0.33	0.30	0.20
EU-SILC(weighted)	0.22	0.28	0.31	0.17

Variable=nfem

	0	1	2	3+
SHIW	0.09	0.59	0.24	0.08
HBS	0.09	0.57	0.25	0.08
EU-SILC	0.11	0.57	0.24	0.08
SHIW(weighted)	0.10	0.57	0.25	0.08
HBS(weighted)	0.11	0.58	0.23	0.07
EU-SILC(weighted)	0.13	0.57	0.23	0.08

Variable=naltrac

	0	1	2	3+
SHIW	0.46	0.26	0.16	0.12
HBS	0.49	0.34	0.12	0.05
EU-SILC	0.52	0.36	0.09	0.02
SHIW(weighted)	0.45	0.25	0.17	0.12
HBS(weighted)	0.52	0.33	0.12	0.04
EU-SILC(weighted)	0.53	0.36	0.09	0.02

Variable=npens

	0	1	2+
SHIW	0.49	0.36	0.15
HBS	0.58	0.30	0.12
EU-SILC	0.61	0.29	0.11
SHIW(weighted)	0.55	0.32	0.13
HBS(weighted)	0.60	0.29	0.11
EU-SILC (weighted)	0.63	0.28	0.09

Variable=ndip

	0	1	2
SHIW	0.50	0.32	0.19
HBS	0.46	0.34	0.20
EU-SILC	0.49	0.33	0.18
SHIW(weighted)	0.47	0.34	0.20
HBS(weighted)	0.47	0.34	0.19
EU-SILC (weighted)	0.49	0.33	0.17

Variable=nindip

	0	1	2
SHIW	0.87	0.11	0.02
HBS	0.80	0.15	0.04
EU-SILC	0.80	0.16	0.04
SHIW(weighted)	0.86	0.11	0.02
HBS(weighted)	0.82	0.15	0.04
EU-SILC (weighted)	0.81	0.17	0.03

Variable=nbam

	0	1	2	3+
SHIW	0.79	0.12	0.07	0.01
HBS	0.78	0.14	0.07	0.01
EU-SILC	0.75	0.14	0.09	0.01
SHIW(weighted)	0.76	0.14	0.08	0.05
HBS(weighted)	0.77	0.14	0.08	0.01
EU-SILC (weighted)	0.75	0.14	0.09	0.01

Variable=tipobam

	0	1
SHIW	0.79	0.21
HBS	0.78	0.22
EU-SILC	0.75	0.25
SHIW(weighted)	0.76	0.24
HBS(weighted)	0.77	0.23
EU-SILC (weighted)	0.75	0.25

Variable=nanz

	0	1	2+
SHIW	0.60	0.26	0.15
HBS	0.61	0.24	0.14
EU-SILC	0.64	0.24	0.12
SHIW(weighted)	0.63	0.24	0.13
HBS(weighted)	0.62	0.25	0.13
EU-SILC (weighted)	0.64	0.24	0.12

Variable=ngiova

	0	1	2+
SHIW	0.52	0.28	0.20
HBS	0.49	0.29	0.22
EU-SILC	0.49	0.29	0.22
SHIW(weighted)	0.47	0.31	0.22
HBS(weighted)	0.51	0.28	0.21
EU-SILC (weighted)	0.49	0.29	0.22

Variable=nadul

	0	1	2+
SHIW	0.41	0.26	0.33
HBS	0.41	0.26	0.33
EU-SILC	0.43	0.27	0.30
SHIW(weighted)	0.43	0.26	0.31
HBS(weighted)	0.44	0.26	0.30
EU-SILC (weighted)	0.46	0.27	0.27

Variable=nmin

	0	1	2	3+
SHIW	0.75	0.14	0.09	0.02
HBS	0.72	0.15	0.10	0.02
EU-SILC	0.71	0.16	0.11	0.02
SHIW(weighted)	0.72	0.15	0.11	0.02
HBS(weighted)	0.73	0.15	0.10	0.02
EU-SILC (weighted)	0.72	0.15	0.11	0.02

Variable=tipoanz

	0	1
SHIW	0.80	0.20
HBS	0.80	0.20
EU-SILC	0.82	0.18
SHIW(weighted)	0.82	0.18
HBS(weighted)	0.81	0.19
EU-SILC (weighted)	0.82	0.19

Variable=nocc

	0	1	2+
SHIW	0.40	0.35	0.27
HBS	0.35	0.35	0.30
EU-SILC	0.38	0.35	0.27
SHIW(weighted)	0.35	0.38	0.28
HBS(weighted)	0.36	0.35	0.28
EU-SILC (weighted)	0.38	0.37	0.26

Variable=mutuo

	0	1
SHIW	0.92	0.08
HBS	0.90	0.098
EU-SILC	-	-
SHIW(weighted)	0.92	0.085
HBS(weighted)	0.90	0.10
EU-SILC (weighted)		

Variable=fitmut

	0	1
SHIW	0.71	0.29
HBS	0.73	0.27
EU-SILC	0.71	0.29
SHIW(weighted)	0.70	0.30
HBS(weighted)	0.72	0.28
EU-SILC (weighted)	0.69	0.30

Variable=tabt

	1	2
SHIW	0.21	0.79
HBS	0.18	0.82
EU-SILC	0.17	0.84
SHIW(weighted)	0.22	0.78
HBS(weighted)	0.19	0.81
EU-SILC (weighted)	0.18	0.82

Variable=ncomp

	1	2	3	4	5+
SHIW	0.24	0.30	0.21	0.19	0.07
HBS	0.23	0.27	0.22	0.21	0.07
EU-SILC	0.25	0.28	0.22	0.19	0.06
SHIW(weighted)	0.25	0.28	0.21	0.19	0.07
HBS(weighted)	0.26	0.27	0.21	0.19	0.06
EU-SILC (weighted)	0.28	0.27	0.21	0.18	0.06

Variable=area

	0	1	2	3+
SHIW	0.25	0.21	0.22	0.33
HBS	0.24	0.20	0.18	0.38
EU-SILC	0.25	0.24	0.24	0.27
SHIW(weighted)	0.27	0.21	0.20	0.32
HBS(weighted)	0.29	0.20	0.20	0.32
EU-SILC (weighted)	0.27	0.21	0.20	0.32

Table A.2 Chi-squared test – Null hypothesis: SHIW and HBS samples are extracted from the same population

Variable	Simple frequencies	Weighted frequencies
ndiploma	X-squared = 14.639 df = 3 p-value = 0.002153	X-squared = 4.3392 df = 3 p-value = 0.2271
nlaurea	X-squared = 0.603 df = 2 p-value = 0.7397	X-squared = 0.1849 df = 2 p-value = 0.9117
nobbligo	X-squared = 6.077 df = 3 p-value = 0.1079	X-squared = 13.9313 df = 3 p-value = 0.003
nfem	X-squared = 12.9384 df = 3 p-value = 0.004772	X-squared = 22.5633 df = 3 p-value = 4.98e-05
naltrc	X-squared = 566.4083 df = 3 p-value < 2.2e-16	X-squared = 1698.51 df = 3 p-value < 2.2e-16
npens	X-squared = 213.654 df = 2 p-value < 2.2e-16	X-squared = 102.9857 df = 2 p-value < 2.2e-16
ndip	X-squared = 25.4916 df = 2 p-value = 2.915e-06	X-squared = 11.2892 df = 2 p-value = 0.003537
nindip	X-squared = 173.3246 df = 2 p-value < 2.2e-16	X-squared = 124.6993 df = 2 p-value < 2.2e-16
nocc	X-squared = 63.2882 df = 2 p-value = 1.808e-14	X-squared = 6.5327 df = 2 p-value = 0.03815
nbam	X-squared = 19.0591 df = 3 p-value = 0.0002658	X-squared = 25.0272 df = 3 p-value = 1.524e-05
tipobam	X-squared = 4.1971 df = 1 p-value = 0.04049	X-squared = 1.9986 df = 1 p-value = 0.1574
nanz	X-squared = 8.1581 df = 2 p-value = 0.01692	X-squared = 4.1845 df = 2 p-value = 0.1234
nadul	X-squared = 1.9238, df = 2 p-value = 0.3822	X-squared = 6.2401 df = 2 p-value = 0.04415
ngiova	X-squared = 21.8905 df = 2 p-value = 1.764e-05	X-squared = 43.3096 df = 2 p-value = 3.940e-10
nmin	X-squared = 17.6236 df = 3 p-value = 0.0005259	X-squared = 19.4816 df = 3 p-value = 0.0002173
tipoanz	X-squared = 0.3779 df = 1 p-value = 0.5387	X-squared = 10.156 df = 1 p-value = 0.001438
mutuo	X-squared = 26.2123 df = 1 p-value = 3.059e-07	X-squared = 18.8076 df = 1 p-value = 1.446e-05
fitmut	X-squared = 7.437 df = 1 p-value = 0.00639	X-squared = 14.27 df = 1 p-value = 0.0001584
tabt	X-squared = 42.5071 df = 1 p-value = 7.043e-11	X-squared = 51.0077 df = 1 p-value = 9.2e-13
ncomp	X-squared = 39.6787 df = 4, p-value = 5.044e-08	X-squared = 20.792 df = 4 p-value = 0.0003482
area	X-squared = 80.9981 df = 3 p-value < 2.2e-16	X-squared = 14.9737 df = 3 p-value = 0.001839

Table A.3 Chi-squared test – Null hypothesis: EU-SILC and HBS samples are extracted from the same population

variable	Simple frequencies	Weighted frequencies
ndiploma	X-squared = 36.5001, df = 3, p-value = 5.87e-08	X-squared = 64124.02, df = 3, p-value < 2.2e-16
nlaurea	X-squared = 25.0797, df = 2, p-value = 3.581e-06	X-squared = 61347.77, df = 2, p-value < 2.2e-16
nobbligo	X-squared = 302.8808, df = 3, p-value < 2.2e-16	X-squared = 241242.6, df = 3, p-value < 2.2e-16
nfem	X-squared = 35, df = 3, p-value = 1.084e-07	X-squared = 33868, df = 3, p-value < 2.2e-16
naltrc	X-squared = 437.0992, df = 3, p-value < 2.2e-16	X-squared = 292554.9, df = 3, p-value < 2.2e-16
npens	X-squared = 63.2701, df = 2, p-value = 1.824e-14	X-squared = 67053.31, df = 2, p-value < 2.2e-16
ndip	X-squared = 58.9569, df = 2, p-value = 1.576e-13	X-squared = 40427.91, df = 2, p-value < 2.2e-16
nindip	X-squared = 13.0975, df = 2, p-value = 0.001432	X-squared = 13744.43, df = 2, p-value < 2.2e-16
nocc	X-squared = 51.2953, df = 2, p-value = 7.267e-12	X-squared = 47468.86, df = 2, p-value < 2.2e-16
nbam	X-squared = 80, df = 3, p-value < 2.2e-16	X-squared = 43402, df = 3, p-value < 2.2e-16
tipobam	X-squared = 63, df = 1, p-value = 2.063e-15	X-squared = 16017, df = 1, p-value < 2.2e-16
nanz	X-squared = 49, df = 2, p-value = 2.563e-11	X-squared = 24985, df = 2, p-value < 2.2e-16
nadul	X-squared = 47, df = 2, p-value = 4.931e-11	X-squared = 36933, df = 2, p-value < 2.2e-16
ngiova	X-squared = 1.6, df = 2, p-value = 0.4551	X-squared = 16161, df = 2, p-value < 2.2e-16
nmin	X-squared = 62, df = 3, p-value = 2.459e-13	X-squared = 22028, df = 3, p-value < 2.2e-16
tipoanz	X-squared = 18, df = 1, p-value = 2.766e-05	X-squared = 2816, df = 1, p-value < 2.2e-16
fitmut	X-squared = 173, df = 1, p-value < 2.2e-16	X-squared = 228397, df = 1, p-value < 2.2e-16
tabt	X-squared = 12, df = 1, p-value = 0.0006531	X-squared = 178, df = 1, p-value < 2.2e-16
ncomp	X-squared = 66, df = 4, p-value = 1.612e-13	X-squared = 25944, df = 4, p-value < 2.2e-16
area	X-squared = 667, df = 3, p-value < 2.2e-16	X-squared = 47, df = 3, p-value = 3.184e-10

Table A.4 Distribution of households according to some characteristic of the Head of the household:
SHIW and HBS data –2004

SEX OF THE HEAD OF THE HOUSEHOLD - % values

	MALE	FEMALE
SHIW	60.97	39.03
HBS	73.31	26.69
SHIW(weighted)	61.19	38.81
HBS(weighted)	71.40	28.60
X-squared = 439.16, df = 1, p-value < 2.2e-16		

AGE OF THE HEAD OF THE HOUSEHOLD -% values

	<30	30<=eta<40	40<=eta<50	50<=eta<65	eta>=65
SHIW	3.23	12.93	18.55	30.34	34.93
HBS	2.00	12.77	19.28	29.32	36.60
SHIW(weighted)	4.12	16.58	21.14	26.63	31.49
HBS(weighted)	2.55	14.07	19.84	27.68	35.83
X-squared = 47.78, df = 4, p-value = 1.049e-09					

EDUCATION ATTAINED OF THE HEAD OF THE HOUSEHOLD (HIGHEST LEVEL ATTAINED) -% values

	up to 8 years' schooling	11-13 years' schooling	University degree or post university studies
SHIW	62.36	29.69	7.95
HBS	63.52	28.25	8.21
SHIW(weighted)	61.42	30.93	7.65
HBS(weighted)	62.60	29.10	8.29
X-squared = 0.0359, df = 2, p-value = 0.9822			

Table A.5 Empirical distribution of households according to some characteristic of the Head of the household:
EU-SILC and HBS data –2004

SEX OF THE HEAD OF THE HOUSEHOLD - % values

	MALE	FEMALE
EU-SILC	71.7	28.3
HBS	73.3	26.7
EU-SILC(weighted)	71.5	28.5
HBS(weighted)	71.4	28.6
X-squared = 15.6, df = 1, p-value = 8.012e-05		

AGE OF THE HEAD OF THE HOUSEHOLD -% values

	<30	30<=eta<40	40<=eta<50	50<=eta<65	eta>=65
EU-SILC	4.11	15.95	18.95	27.46	33.54
HBS	2.01	12.77	19.28	29.33	36.61
EU-SILC(weighted)	4.07	16.01	19.14	27.66	33.12
HBS(weighted)	2.56	14.07	19.85	27.69	35.83
X-squared = 298, df = 4, p-value < 2.2e-16					

EDUCATION OF THE HEAD OF THE HOUSEHOLD (HIGHEST LEVEL ATTAINED) - % values

	up to 8 years' schooling	9-13 years' schooling	University degree or post university studies
EU-SILC	61.52	29.45	9.03
HBS	63.52	28.25	8.21
EU-SILC(weighted)	61.21	29.56	9.22
HBS(weighted)	62.60	29.10	8.29
X-squared = 5.74, df = 2, p-value = 0.05672			

Since the definitions of head of the household cannot be harmonized on the basis of the information supplied by the surveys, we have decided not to include such variables in the matching process.
On the basis of previous analysis we have decided to exclude also the variables which present very different distributions, particularly **altrac** and **npens**. For this reason, these variables will not be considered in the analysis which follow.

A.2 Selecting the matching variables

The analysis carried out here aims at identifying the subset of variables that best explain income and consumption. We focus on the subset of variables whose distributions are considered comparable on the basis of previous analysis.

Various techniques can be used (see D'Orazio et al 2006).

The Analysis of variance (ANOVA)

As a first step we have carried out an ANOVA. Income and consumption are considered as dependent (or response) variables[12], the harmonized variables are the factors. We test the hypothesis that income (consumption) mean is the same among groups of household identified by different levels of the factor variable. Consider for example variable Ncomp (number of members of the household) as the factor variable: we test the hypothesis that income (consumption) means are equal for households groups regardless of the household dimension (1, 2, 3 or more members).

According to Test F, the null hypothesis must be rejected for each factor for both income and consumption expenditure. The only exception is variable FITMUT for which the null hypothesis cannot be rejected with $\alpha=0.05$. Almost all the variables indeed, seem to effect household income and consumption levels.

The ETA squared coefficient ($\eta 2$) supplies a measure of the strength of the link between Income (or consumption expenditure) and the considered factor. The coefficient can be interpreted as the portion of deviance of the dependent variable explained by the factor.

Table A.6 Eta squared ($\eta 2$) coefficients

response = log(Y) -SHIW04-		response = log(C) -SHIW04-		response= log(C) -hbs04-	
	η^2		η^2		η^2
qalim	0.31	qalim	0.40	qalim	0.48
nocc	0.26	nocc	0.22	ncomp	0.23
ncomp	0.18	ncomp	0.18	nocc	0.22
supab	0.18	ndiploma	0.18	ndiploma	0.13
ndiploma	0.17	supab	0.18	nadul	0.13
ndip	0.15	ndip	0.13	ndip	0.13
nlaurea	0.13	nadul	0.12	nanz	0.12
nadul	0.11	nlaurea	0.11	ngiova	0.11
tabt	0.10	nanz	0.09	tipoanz	0.10
area	0.08	ngiova	0.08	nfem	0.09
ngiova	0.07	area	0.07	nmin	0.06
nanz	0.05	nobbligo	0.06	nlaurea	0.06
nobbligo	0.05	nfem	0.06	nindip	0.05
nfem	0.05	tipoanz	0.05	nobbligo	0.05
nindip	0.04	tabt	0.05	nbam	0.04
fitmut	0.04	nindip	0.04	tipobam	0.04
tipoanz	0.03	nmin	0.04	mutuoab	0.03
mutuoab	0.02	nbam	0.03	area	0.03
nmin	0.02	tipobam	0.03	tabt	0.01
nbam	0.01	mutuoab	0.02	fitmut*	0.00
tipobam	0.01	fitmut	0.01	* P-VALUE=0.052	

Variables are ordered from the one with the greatest eta square value down to the smallest.

[12] From here onwards we consider the logarithmic transform both to income and consumption expenditure

The quintile of consumption expenditure (Qalim) is the variable which explains the largest portion of deviance both for income and consumption expenditure. This variable is followed by the number of members (Ncomp) and the number of employed (Nocc).

Among the education related variables, Ndiploma and Nlaurea turn out to be the most strongly connected with income and consumption expenditure.

Among the age related variables, Nadul, Nanz and Ngiova have the largest Eta square values.

The number of employees variable (Ndip) explains a larger portion both of income and consumption expenditure (around 0.15% for income and 13% for consumption) with respect to the number of self-employed (Nindip).

Geographical area of residence (Area) shows an intermediate ranking value of η2 with respect to SHIW income and consumption expenditure. Surprisingly Area is one of the weakest variables in explaining consumption expenditure according to HBS data.

In the end it is worth stressing the weak explanatory power of house related variables with the exception of the house surface variable (Supab). Unfortunately, such variable is not disposable in the Istat survey microdata set.

Multiple regression models

As a second step, we have estimated multiple regression models with income (consumption expenditure) as dependent variable and the whole subset of harmonized variables as dependent variables.

By means of a stepwise regression[13], we have selected the subset of independent variables that best explain the average income and consumption expenditure. These are the results:

MODEL A- dependent variable: SHIW income
The stepwise regression procedure has excluded the following regressors: Tabt, Ngiova, Tipobam.
Multiple R-squared: 0.6077, Adjusted R-squared: 0.606

MODEL B – dependent variable= SHIW consumption expenditure
The stepwise regression procedure has excluded the following regressors: Tabt, Nmin, Nindip, Tipobam
Multiple R-squared: 0.4683, Adjusted R-squared: 0.4665

MODEL C – dependent variable= HBS consumption expenditure
The stepwise regression procedure has excluded the following regressors: Nmin, Ndip, Tipobam, Ngiova.
Multiple R-squared: 0.3916, Adjusted R-squared: 0.3909

The following tables show the coefficients of the models.

[13]We have applied the R function step based on the AIC criterion

Tab. A.7 - MODEL A- dependent variable: SHIW income

Coefficients	Estimates	Std.Error	t value	Pr(>|t|)	
(Intercept)	9.485	0.024	400.755	< 2e-16	***
as.factor(NLAUREA)1	0.303	0.016	19.227	< 2e-16	***
as.factor(NLAUREA)2	0.550	0.026	21.008	< 2e-16	***
as.factor(NANZ)1	0.216	0.017	12.741	< 2e-16	***
as.factor(NANZ)2	0.362	0.024	15.027	< 2e-16	***
as.factor(NADUL)1	0.071	0.014	5.171	2.39E-07	***
as.factor(NADUL)2	0.143	0.018	8.048	9.61E-16	***
as.factor(NOCC)1	0.423	0.027	15.517	< 2e-16	***
as.factor(NOCC)2	0.804	0.045	17.836	< 2e-16	***
as.factor(NDIP)1	-0.137	0.025	-5.405	6.68E-08	***
as.factor(NDIP)2	-0.239	0.043	-5.559	2.80E-08	***
as.factor(NINDIP)1	-0.050	0.025	-1.988	0.046811	*
as.factor(NINDIP)2	-0.096	0.050	-1.909	0.056304	
as.factor(FITMUT)1	-0.384	0.012	-31.809	< 2e-16	***
as.factor(NCOMPr)2	0.133	0.018	7.221	5.62E-13	***
as.factor(NCOMPr)3	0.102	0.024	4.258	2.09E-05	***
as.factor(NCOMPr)4	0.099	0.030	3.327	0.000882	***
as.factor(NCOMPr)5	0.077	0.039	1.961	0.049892	*
as.factor(AREA)2	-0.007	0.014	-0.499	0.618043	
as.factor(AREA)3	-0.048	0.014	-3.461	0.000541	***
as.factor(AREA)4	-0.306	0.013	-23.833	< 2e-16	***
as.factor(QALIM)2	0.179	0.016	10.889	< 2e-16	***
as.factor(QALIM)3	0.275	0.017	16.067	< 2e-16	***
as.factor(QALIM)4	0.369	0.017	22.115	< 2e-16	***
as.factor(QALIM)5	0.508	0.019	27.441	< 2e-16	***
as.factor(NFEM)1	-0.130	0.019	-6.814	1.02E-11	***
as.factor(NFEM)2	-0.162	0.023	-6.941	4.19E-12	***
as.factor(NFEM)3	-0.164	0.029	-5.623	1.94E-08	***
as.factor(TIPOANZ)1	-0.037	0.015	-2.452	0.014218	*
as.factor(NMIN)1	-0.044	0.017	-2.550	0.01079	*
as.factor(NMIN)2	-0.067	0.024	-2.803	0.005077	**
as.factor(NMIN)3	-0.045	0.045	-1.011	0.312029	
as.factor(MUTUOAB)1	0.359	0.020	17.934	< 2e-16	***
as.factor(NDIPLOMA)1	1 0.14667(	0.012	12.452	< 2e-16	***
as.factor(NDIPLOMA)2	2 0.21529:	0.016	13.518	< 2e-16	***
as.factor(NDIPLOMA)3	3 0.27800:	0.027	10.300	< 2e-16	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Tab. A.8 - MODEL B – dependent variable= SHIW consumption expenditure

Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.52264	0.02218	429.296	< 2e-16	***
as.factor(NLAUREA)1	0.27999	0.01502	18.645	< 2e-16	***
as.factor(NLAUREA)2	0.52898	0.02464	21.471	< 2e-16	***
as.factor(NANZ)1	0.03552	0.01608	2.21	0.02716	*
as.factor(NANZ)2	0.09719	0.02441	3.981	6.92E-05	***
as.factor(NADUL)1	0.01393	0.01503	0.927	3.54E-01	
as.factor(NADUL)2	0.0578	0.01989	2.907	3.66E-03	**
as.factor(NOCC)1	0.27061	0.01909	14.179	< 2e-16	***
as.factor(NOCC)2	0.44176	0.02505	17.634	< 2e-16	***
as.factor(NDIP)1	-0.12275	0.01656	-7.413	1.36E-13	***
as.factor(NDIP)2	-0.15477	0.02278	-6.794	1.17E-11	***
as.factor(FITMUT)1	-0.1981	0.01165	-17.007	< 2e-16	***
as.factor(NCOMPr)2	0.24397	0.01815	13.44	< 2e-16	***
as.factor(NCOMPr)3	0.32086	0.0244	13.149	< 2e-16	***
as.factor(NCOMPr)4	0.35046	0.02845	12.319	< 2e-16	***
as.factor(NCOMPr)5	0.42631	0.03508	12.151	< 2e-16	***
as.factor(AREA)2	0.03577	0.01341	2.667	0.00766	**
as.factor(AREA)3	0.06505	0.0133	4.89	1.03E-06	***
as.factor(AREA)4	-0.24519	0.01235	-19.848	< 2e-16	***
as.factor(NFEM)1	-0.08348	0.01849	-4.514	6.45E-06	***
as.factor(NFEM)2	-0.12947	0.02261	-5.726	1.07E-08	***
as.factor(NFEM)3	-0.16377	0.02826	-5.794	7.12E-09	***
as.factor(TIPOANZ)1	-0.06172	0.01448	-4.264	2.03E-05	***
as.factor(NGIOVA)1	-0.04156	0.01522	-2.73	0.00635	**
as.factor(NGIOVA)2	-0.10287	0.02222	-4.629	3.73E-06	***
as.factor(MUTUOAB)1	0.22477	0.01936	11.608	< 2e-16	***
as.factor(NDIPLOMA)1	0.17488	0.01138	15.37	< 2e-16	***
as.factor(NDIPLOMA)2	0.26965	0.01534	17.573	< 2e-16	***
as.factor(NDIPLOMA)3	0.37785	0.02581	14.642	< 2e-16	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Tab. A.9 - MODEL C – dependent variable= HBS consumption expenditure

Coefficients:					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.518185	0.081498	116.791	< 2e-16	***
as.factor(NLAURBF)1	0.254726	0.012526	20.335	< 2e-16	***
as.factor(NLAURBF)2	0.425073	0.020462	20.773	< 2e-16	***
as.factor(ANZBF)1	-0.041486	0.01271	-3.264	0.0011	**
as.factor(ANZBF)2	0.058208	0.01743	3.34	0.00084	***
as.factor(ADULBF)1	0.089112	0.010557	8.441	< 2e-16	***
as.factor(ADULBF)2	0.113638	0.012208	9.309	< 2e-16	***
as.factor(NOCCBF)1	0.143203	0.012274	11.667	< 2e-16	***
as.factor(NOCCBF)2	0.215833	0.01507	14.322	< 2e-16	***
as.factor(NINDIPBF)1	0.076247	0.011298	6.749	1.53E-11	***
as.factor(NINDIPBF)2	0.080115	0.02084	3.844	0.000121	***
as.factor(FITMUTBF)1	-0.312007	0.080263	-3.887	0.000102	***
as.factor(NCOMPBF)2	0.370954	0.014227	26.073	< 2e-16	***
as.factor(NCOMPBF)3	0.487266	0.017047	28.584	< 2e-16	***
as.factor(NCOMPBF)4	0.553074	0.019309	28.644	< 2e-16	***
as.factor(NCOMPBF)5	0.683169	0.023519	29.047	< 2e-16	***
as.factor(AREABF)2	-0.046477	0.01157	-4.017	5.91E-05	***
as.factor(AREABF)3	-0.138471	0.011832	-11.703	< 2e-16	***
as.factor(AREABF)4	-0.343819	0.01024	-33.576	< 2e-16	***
as.factor(NFEMMBF)1	-0.078396	0.015482	-5.064	4.14E-07	***
as.factor(NFEMMBF)2	-0.069406	0.018647	-3.722	0.000198	***
as.factor(NFEMMBF)3	-0.092791	0.022811	-4.068	4.76E-05	***
as.factor(TANZBF)1	-0.199958	0.012639	-15.821	< 2e-16	***
as.factor(MUTUOBF)1	0.431342	0.07923	5.444	5.25E-08	***
as.factor(NDIPLBF)1	0.181285	0.009501	19.081	< 2e-16	***
as.factor(NDIPLBF)2	0.229859	0.01232	18.657	< 2e-16	***
as.factor(NDIPLBF)3	0.312891	0.019779	15.82	< 2e-16	***
as.factor(TABBF)2	-0.141281	0.079594	-1.775	0.075907	.
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Cramer 's V coefficient

Cramer's V is a statistic measuring the strength of association or dependency between two (nominal) categorical variables in a contingency table.

We calculate Cramer's V to assess the association of income (consumption expenditure) and the harmonized variables.

Income and consumption expenditure have been categorized into 8 quantiles.

Table A.10 shows the coefficient values.

The range of V is (0,1): the closer V is to 0, the smaller the association between the categorical variables.

Tab. A.10 Cramer's V coefficients

data source: shiw			data source: shiw			data source: hbs04	
	Cramer'V (income)			Cramer'V (consumption)			Cramer'V (consumption)
nocc	0.396		qalim	0.371		qalim	0.43
qalim	0.324		nocc	0.359		nocc	0.35
ndip	0.320		ndip	0.279		tipoanz	0.34
tabt	0.319		ndiploma	0.263		ndip	0.27
nlaurea	0.293		nadul	0.255		nadul	0.27
ndiploma	0.271		nlaurea	0.247		nanz	0.26
nadul	0.267		tipoanz	0.239		ncomp	0.25
ncomp	0.252		nanz	0.236		ngiova	0.24
nanz	0.208		ncomp	0.233		ndiploma	0.22
ngiova	0.203		tabt	0.217		tipobam	0.22
fitmut	0.202		ngiova	0.210		mutuoab	0.18
tipoanz	0.190		nobbligo	0.179		nfem	0.18
nobbligo	0.179		tipobam	0.179		nlaurea	0.17
area	0.177		area	0.167		nindip	0.16
nindip	0.154		mutuoab	0.163		nobbligo	0.15
mutuoab	0.154		nindip	0.148		nmin	0.15
nfem	0.146		nfem	0.144		tabt	0.13
tipobam	0.111		nmin	0.125		nbam	0.13
nmin	0.090		nbam	0.107		area	0.10
nbam	0.076		fitmut	0.104		fitmut	0.01

The Cramer's V values roughly confirm previous analysis. Only variable TABT (house renting) seems to diverge. This variable records one of the highest Cramer's V values with respect to income quintiles whereas it results as one of the excluded regressor in the multiple regression model.

Appendix B - Disposable income in NA and in the income surveys: differences in definitions of income components

- Consumer household *gross operating surplus* as estimated in Italian NA derives from the following activityes::
  - o own account production of housing services by owner occupied (imputed rents),
  - o own-account production of agricultural products,
  - o own- account manteinance of dwellings.

  Of these kind of activities, surveys only estimate imputed rents (which count for the greatest part of gross operating surplus). There is information about mortgages payback, but it is impossible to compute the Fisim paid on interests, since there is no information to separate the amount paid in terms of interest from the one paid as capital. While EU-SILC asses imputed rents only on households residence, SHIW compute as well imputed rents on other dwellings at households disposal.
- *Share of mixed income from producer households*. Istat estimates this amount deducting land rent, current taxes and interests paid by producer households, from net mixed income of producer households. It contains as well production coming from underground economy. Note that producer households mixed income includes not only the gross operating surplus coming from the market production activity of units classified in the productive households sector, but also rents received by households, considered as output from market activities, therefore added to producer households mixed income. From SHIW it is possible to estimate mixed income, since it reports both the necessary monetary variable and information about business legal status and number of workers (not number of employees, some approximation is needed). SHIW reports as well income received from rents of owned dwellings (separable from land rents). Since EU-SILC does not ask information about legal status of the firm where the self employed works it is not possible to separate mixed income from producer households from other self employment incomes. To compare surveys and NA estimates we deducted from NA mixed income part of social contribution paid by self employed (D6113), computed as quote of registered self employed of producer households over total of registered self employed. NA figure still includes all current taxation.
- *Compensation of employees D1*: in National accounts it includes wages and salaries (D11) and employers' social contributions (D12), actual and imputed, gross of taxation (D5) and employees social contributions (D6112) that, in sector accounts, is deducted in the secondary distribution of income. D1 includes as well income of non registered workers, which is hardly declared in surveys. SHIW and EU-SILC ask questions only about wages and salaries net of taxations and employees social contributions. Wages and salaries have the same definitions in surveys and in National accounts (ESA95 §4.03-4.05), also wages and salaries in kind are included. Net wages and salaries are reported for primary and secondary positions, for all households components. Therefore to compare NA and survey figures only wages and salaries have to be considered; employers' social contribution (D12) is not considered, but , at the same time, not deducted in secondary distribution of income (D612). Once NA employees social contributions

(D6112) is deducted from NA wages and salaries (D11), NA and survey figures should differ only for taxation and underground economy.

- Property income (D4): It "is the income receivable by the owner of a financial asset or a tangible non produced asset in return from providing funds or putting the tangible non –produced asset at the disposal of another institutional unit" (ESA95 §4.41). It includes
  - *Interests* (D41). Surveys computes interests (paid and received) by multiplying the appropriate interest rate by each financial instrument held by the household. NA and surveys consider the same financial instrument, but SHIW computes interests including in financial stocks also shares, it is therefore necessary to separate shares from the other financial asset before re-computing interests. Interest rate are the appropriate[14] average rates of return for each financial asset. NA report paid and received interest corrected and not corrected for Fisim. To compare figures and to approach to survey definition we need to display interest not corrected for Fisim[15]. Comparability problem derive not from difference in definition or computation since it is possible to re-compute interests to match up to NA definitions (i.e. excluding shares from financial assets), but rather for survey underreporting or selection bias (which bring as well to under-evaluation of Fisim).
  - Distributed income from corporations (D42). This item includes dividends, which can be considered a "pure" property income, and withdrawals from quasi-corporations (D422) and from corporations (D423). This last category of property income (D423) has been introduced by Italian Sector accounting, since in Italy legal status of corporations does not necessarily implies net separation between labour and capital, there is a considerable number of self employed working in Corporations sector (S11) but not in quasi corporations (only 66,4% of S11 self employed work in quasi-corporations). Since Istat compute a full set of accounts up to gross operating surplus for all sub sector of non financial corporations and financial corporations defined by legal status and size, it was decided to remunerate all these self employed with net disposable income of the sub-sector in which they work using appropriate indicators to decompose each flow from gross operating surplus to net disposable income among corporations sub-sectors: quasi corporations, other (small) corporations with self employed.
    - *Dividends* (D421) "are a form of property income received by owner of shares". SHIW does not ask amount received in form of dividends, but compute it, together with interests, as stock multiplied by rate of returns. It is therefore possible to separate them. EU-SILC reports dividends together with other property incomes, it is not possible to separate them.
    - Withdrawals from quasi-corporations (D422): it is possible to compute the corresponding SHIW self employed income since there are questions about company legal status and class size. NA aggregate includes income from underground economy, social contribution paid by self employed and taxation. To compare SHIW and NA estimates we deducted from NA D422 part of social contribution paid by self

[14] Survey is carried out by Italian central bank, they have all information about financial market.

[15] That is why NA net disposable income reported in table 3 is not the amount reported in sector accounts tables, but is re-computed considering interests non corrected for Fisim.

employed (D6113), computed as quote of registered self employed of producer households over total of registered self employed. It is not possible to compute the corresponding EU-SILC value, since the legal status is not reported.

- Withdrawals from corporations (D423) it is possible to compute the corresponding SHIW self employed income since there are questions about company legal status and class size. NA aggregate includes income from underground economy, social contribution paid by self employed and taxation. To compare SHIW and NA estimates we deducted from NA D423 part of social contribution paid by self employed (D6113), computed as quote of registered self employed of producer households over total of registered self employed. It is not possible to compute the corresponding EU-SILC value, since the legal status is not reported.
- Other property income: It includes *property income attributed to insurance policy holders* (D44): it is an imputed flow from insurance companies to households to attribute them income received by insurance company and pension funds from the investment of insurance technical reserves. It is not computed nor computable in surveys. *Rents on land* (D45) SHIW reports rents received by households of each type of fixed asset it is therefore possible separate rents on land from other rents. EU-SILC, instead, reports rents on land together with other property incomes, it is not possible to separate them.

- Social contributions (D61): Surveys does not report social contributions. To compare surveys with NA labour income, actual (D6111) and imputed (D612) employers' social contribution, employees social contributions (D6112), have been deducted from NA wages and salaries displayed in tables 3a and 3b. Therefore it does not have to be deducted to get disposable income. At the same way self employed social contribution (D6113) has been deducted proportionally from each typology of NA self employed incomes.
- Social benefits (D62): "social benefits are transfers to households, in cash or in kind, intended to relieve them from the financial burden of a number of risks or needs, made through collectively organized schemes" (ESA95 4.83). Also surveys reports social benefits received by households, detailing by kind of benefits and paying board. SHIW classifies social benefits in two items that have therefore to be summed up. The first item records incomes received from social security boards by persons classified as pensioners. The second item records income classified in the survey as "transfers/other income" but that are due in case of risks or needs that ESA95 (§4.84) classifies as giving rise to social benefits (for example unemployment benefits). EU-SILC, instead, records social benefits according to the ESA95 classification of risks and needs, therefore this aggregate is coherent with NA.
- NA Current transfers include:
  - D71, non life insurance premiums paid by households. Surveys on income reports only incomes, not expenditure, but SHIW has a special section on insurance, since it is a form of financial saving. It reports non life insurance premium paid by households, except motor third party liability, therefore this aggregate is not totally comparable.

- D72, non life insurance claims. Fully comparable in NA and SHIW. EU-SILC does not have this information.
- Miscellaneous current transfers (D75). They include, among D75 paid current transfers to Npish, fines and penalties, transfers to other households, which are not reported in surveys, since considered as expenditure; transfers to/from the rest of the world (remittances from migrants) which are hardly included since it is unlikely that resident migrants are surveyed. both surveys record instead transfers from Npish (such as scholarship) and transfers from other households.

Total net NA and surveys current transfers therefore are not fully comparable.

- Current taxation (D5). For ESA 95 (4.77) "current taxes on income wealth etc cover all compulsory, unrequited payments, in cash of in kind, levied periodically by general government and by the rest of the world on the income and wealth…". As said before, surveys consider all monetary values net of taxation.