

Developing Households' sub-sectors accounts: Pros and cons of the top-down and the bottom-up methods

Alessandra Coli^{a,*} and Francesca Tartamella^b

^a*Department of Economics and Management, Pisa, Italy*

^b*Italian National Institute of Statistics (Istat), Roma, Italy*

Abstract. The last decade debate on progress and well-being has stressed the necessity of putting people first, also within the national accounts (NAs) system. The sub-sectoring of Households accounts by groups of households would certainly represent a way of pursuing such objective. Two methods can be used: the top-down method, which breaks down NAs totals (top) according to a set of indicators derived from micro data on household economic resources; the bottom-up method, which uses micro data (bottom) to derive NAs totals. In this paper, we discuss pros and cons of the two methods especially from a practical point of view. Particularly, we present an application of the top-down method for Italy and describe the on-going process towards a full-fledged bottom-up approach for the building of Italian National Accounts.

Keywords: National accounts, Households' sub-sectors accounts, macro-micro reconciliation

1. Introduction

National Accounts (NAs) are based on the transactor/transaction approach, according to which the economy is divided into sectors representing groups of transactors and the transactions of sectors are recorded in accounts. In time, national accountants have answered the increasing demand of information by adding details in the accounting system. On the one hand, they have detailed further the accounts thus singling out very specific kinds of transactions (e.g. those relating to the distributive processes or the analyses by industry and products). On the other hand, they have disentangled transactions by finer homogeneous groups of transactors, thus developing sector and sub-sectors accounts. The result of this double-sectoring process is a very rich source of information, which however gives a significant lower attention to households (i.e. people) and their accounting, than to processes and market transaction flows. With respect to the other institutional sectors, in various countries,

Households is treated as a sort of “residual” sector for most transactions, still far from being homogeneous. Furthermore, NAs limit to describe the economic behavior of a representative household, without any further insight into the heterogeneity of the sector.

The last decade debate on progress and well-being has stressed the necessity of putting people first, also within the NAs system. The development of sub-sectors accounts certainly would help pursuing such objective. Two methods could be used: the top-down method, which breaks down NAs totals (top) according to a set of indicators derived from households' microdata; the bottom-up method, which considers micro data on household economic resources (bottom) as an input to derive NAs totals. Whatever the chosen approach, the reconciliation of micro and macro data on households' accounting becomes crucial. The challenge of national accountants is to find the best possible way to reconcile or align them.

The paper focuses on the theoretical and practical aspects connected with the combination of macro and micro data for the building of Households' sub-sectors accounts. Section 2 provides a quick overview on macro and micro official statistics on Households' accounting. Section 3 discusses the main methodologi-

*Corresponding author: Alessandra Coli, Department of Economics and Management, via Cosimo Ridolfi 10, 56124 Pisa, Italy. E-mail: alessandra.coli1@unipi.it.

cal aspects connected with the building of Households' sub-sectors accounts. Section 4 focuses on the top-down approach, showing an application for Italy based on the statistical matching of income and consumption micro data from two sample surveys. Section 5 describes the ongoing process of micro data integration at the Italian National Institute of Statistics (Istat), which in the next future could lead to bottom-up estimates of Households' accounts. Finally, Section 6 provides some comments and remarks.

2. Data on Households' accounting in official statistics: The macro and micro perspective

In official statistics, two different kinds of data sources provide information on households' economic resources. On the one hand, NAs provide measures for the country as a whole and for groupings of transactors (sectors), there including the Households sector. On the other hand, micro sources (surveys, administrative records, censuses) supply information on earnings and outlays of individuals and households. Both the approaches have limits and strongpoints.

Micro data allow one to analyse the distributions of income, consumption and wealth across the population, for various subgroups, and over time. The collection of such data is essential for studying poverty and its incidence on different socio-economic groups within society. The Canberra Group Handbook on Household Income Statistics [1] represents the international guide for producers and users of micro data on household income. Nevertheless, micro sources typically fail to record all types of economic resources. For example, they do not directly measure the value of social transfers in kind received by households (although they provide information that is useful for imputing them, based on various assumptions), they may fail to capture part of income from non-observed economy (although they provide valuable information to detect non regular labour units) and suffer for problems of under-reporting and non-reporting. These are important limits, especially when the purpose is that of comparing countries with different welfare systems and degree of underground economy. Furthermore, micro data on income, consumption and wealth are seldom coherent, unless they all come from one single data source.

On the other hand, NAs provide an overview of all economic processes, recording how production is generated, how income originated in production flows to

the economic agents and how the economic agents allocate income to consumption, saving and investment. Balance sheet accounts record all the assets and liabilities of each institutional sector, thus providing a measure of their total wealth at a given date. The estimate of aggregates and flows stem from the comparison of different data sources and this method allows pointing out each data source's deficiencies. One major advantage of NAs is the consistency and coherence of households' accounting data on income, consumption and wealth, which is a very desirable property when the purpose is to analyse households' material well-being [2]. Furthermore, NAs allow one to analyze households' economic behavior *vis à vis* the other sectors, given the consistency and coherence of the overall system. However, NAs are not adequate for supporting the design and evaluation of policies aimed at improving the distribution of income or at reducing discrimination since they limit to analyze the economic behavior of a representative household (not considering the problem of aggregation over individuals [3]). This depends on the fact that official statistics institutions never put households' accounting among priorities. Also nowadays, the EU Transmission program¹ envisages only a territorial sub-sectoring of households accounts. This disregard towards households as economic agent is mirrored in the NAs estimate methods, which are mostly based on businesses or governments' records also when estimating transactions primarily involving households. This limit particularly affects distributive analysis where households are usually pictured and estimated as the counterpart sector of distributive transactions.

In short, macro data assure comprehensiveness, coherence and international comparability but cannot provide information on the distribution of economic resources and uses among people; while micro data allow calculating distribution indicators but do not cover all people's economic resources and may give incoherent estimates on interlinked phenomena (e.g. consumption expenditure *versus* disposable income) when data come from independent data sources.

Over the years, the macro and micro approaches have developed separately, sometimes leading to divergent results. In order to bridge macro and micro per-

¹ Programme of National Accounts data delivery within the framework of the new European System of National and Regional Accounts (ESA 2010), as defined in Annex B of the Council Regulation (EU) No 549/2013 of the European Parliament and of the Council of 21 May 2013.

spectives, economists are used to read NAs average values jointly with distribution indicators from micro data sources. However, this is not always correct in that NAs aggregates may not fit the micro aggregates for which distribution indicators are calculated. For example, income distribution indexes are not appropriate to describe the distribution of NAs adjusted disposable income, since they do not take into account the part of income from hidden economy that surveys fail to capture, benefits in kind received from governments or other imputed transactions included in the macro aggregate.

Ideally, NAs measures should incorporate distribution information on households' accounting, since they offer the ideal framework where to fit data coming from different data sources, both establishing the basic control totals and relating the various parts of the system to one another [4]. Performing distributive analyses within the NAs framework (instead of using micro data) would allow one to link households income distribution directly to the performance of the productive system as well as to the economic behavior of all the other economic operators.

NAs directives have repeatedly suggested bringing distribution indicators within Households' accounts through the grouping of households according to economic, socio-economic or geographic criteria or by encouraging the compilation of social accounting matrices [5–7].

In 2009, the SSF Commission on the measurement of economic performance and social progress renewed attention on the topic [2]. The Commission recommends to give prominence to the distribution of income, consumption and wealth ([2], Recommendation 4, p. 40), adding how “ideally distributional measures should be compatible in scope with average measures from the national accounts” ([2], §43, p. 34). The SSF report suggests developing distributional measures of full income, which implies distributing market income but also imputed income such as imputed rents from own-occupied housing, and government services provided in kind ([2], §57, p. 39). NAs' adjusted disposable income is the income measure which best fits the concept of full income recalled by the SSF report.

More recently, Eurostat, OECD and other bodies have been taking significant steps in order to address the Stiglitz Commission recommendations. Among these, the OECD-Eurostat Expert Group on Disparities in a National Accounts framework was devoted to deepen research on the reconciliation of macro (national accounts) and micro (household surveys) esti-

mates with the ultimate purpose of estimating some significant NAs items by households sub-sectors. M. Fessau et al. [8,9] present the main results of the Expert Group whereas Fessau and Van de Ven [10] present some inequality measures on the distribution of income based on such results.

3. Building Household sub-sectors accounts: Theoretical aspects

Building Households' sub-sectors implies combining information from different data sources without losing the consistency of the overall accounting system. This, indeed, is the problem that national accountants have to face currently.

Households' sub-sectors accounts can be built following two different approaches:

- Top-down approach: NAs items are considered as the standard to be broken down using information from individual-based data [9]. This approach is based on the assumption that NAs' estimates of aggregate are more reliable than those coming from micro data sources. This is a common assumption indeed, e.g. researches aimed at assessing the quality of sample surveys or at validating input data for micro-simulation models usually consider NAs' totals as an external yardstick.
- Bottom-up approach: micro data on households' economic resources are directly used to estimate Households' sub-sectors accounts. The underlying assumption is that the divergences between existing micro and macro data must be bridged meeting somewhere in the middle [11]. Micro-macro reconciliation should then work both ways, in the awareness that micro data can contribute to improving the quality and consistency of the NAs estimates [4].

The bottom-up approach is quite demanding but it guarantees the inner coherency and better accuracy of results. Conversely, the top down approach is less difficult and can be applied without changing NAs methodology, at the expense of a probable minor accuracy of final estimates. Whatever the chosen approach, the use of micro data from several sources is essential in that no single micro data set contains the information required for estimating all NAs variables by households' sub-groups [12–14]. Independent data sources often provide conflicting estimates on the same phenomenon or inconsistent estimates on inter-connected phenomena (e.g. income and consumption expenditure). Thus,

the synthesis of micro data from several sources is essential to produce consistent information on the economic behavior of different households' typologies.

This issue of discussion is not new. More than forty years ago, Nancy Ruggles and Richard Ruggles suggested using record linkage and statistical matching techniques in order to match data from different sources, whereas they discarded the use of imputation by regression when the objective, like in NAs, is that of transferring complex sets of information from one source to another [14].

Record linkage is a technique, which compares records contained in two files *A* and *B*, in order to determine pairs of records referred to the same population unit. The identification is based on unit identifiers supposed to be free of errors (e.g. Personal Identification Number, VAT code, etc.). Record linkage between two files is very simple if each record in both files contains the same identifier and this identifier is supposed to be free of errors. In this case, the problem is solved by simply picking out the records (if any) with the same identifier value. Unfortunately, some errors may occur because incorrect information is obtained from the individual, or because information is incorrectly recorded. Due to such errors, two records for the same person may not agree, and two records that agree may refer to different persons. Formalizing the linking procedure into a statistical model, it is possible to measure the probability of generating false-matched-pairs and false-unmatched pairs [15]. In deterministic record linkage, the decision whether a pair of records refers to the same entity or not is based on *ad hoc* rules whereas in probabilistic record linkage the decision is based on statistical models.

Assuming not to have confidentiality problems, one could use record linkage techniques to retrieve information on the same individual/household from census records, taxes records or social security records.

Record linkage can be used when the data sources to be matched contain the same entity i.e. when the linkage involves one sample survey and one administrative data source or two administrative data sources covering the same population. For example, record linkage would allow one to combine information from a sample survey on households' consumption expenditure with data on income coming from tax records.

Conversely, it is not possible to run record linkage when two sample surveys are involved, since the probability of the same individual appearing in both may be very small. This happens, for example, when there is the need of getting households income and consump-

tion expenditure information from two different sample surveys [16]. In this context, the use of statistical matching seems the more appropriate choice. In fact, statistical matching is used to link independent samples of data *A* and *B*, by means of some variables common to both data files [17,18]. Suppose some variables *Y* appear only in *A*, some variables *X* appear only in *B* and a set of variables *Z* can be observed in both samples. Statistical matching generates an artificial data set where each unit records *Z*, *Y* and *X* values.

In order to apply statistical matching it is necessary that:

- *A* and *B* contain both a set of common variables *Z* and a set of specific variables *Y* and *X* respectively;
- the units observed in *A* and *B* have been drawn independently from the same population, so that the number of identical individuals in both datasets is typically small if not zero.

The key assumption of statistical matching is that the set of variables, *Y*, observed in *A* but not in *B*, and the set of variables *X*, observed in *B* but not in *A*, is independent conditionally on the set of common variables *X* (see especially [17–19]). Conditional independence assumption (CIA) is a strong constraint to the application of traditional statistical matching, to the point that a debate is born on the pros and cons of statistical matching (see [18] for a synthesis). Sceptics assert that statistical matching does not bring any additional information on the relationship between the not jointly observed variables. 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 [14]). According to this viewpoint, CIA can be roughly satisfied by carefully selecting the *Z* common variables. According to other scholars, CIA can be satisfied if some additional auxiliary information can be exploited in the matching application. For major details on the usage of auxiliary information in the statistical matching see [17,20].

4. The top-down approach: An application to Italian NAs using statistical matching

In building Households' sub-sectors account, the bridge between the secondary distribution of income account and the Use of income account is quite problematic. In fact, data supporting the sub-sectoring of income items usually differ from those used to breakdown consumption items. On the other hand, the link

Table 1
Average consumption propensities by geographical area – Italy, 2004

Data source	North-West	North-East	Centre	South	Italy
Shiw	72.93%	71.54%	75.52%	81.08%	75.09%
Hbs-Shiw	100.37%	87.76%	82.06%	107.11%	94.79%
Hbs-ItSilc	86.31%	84.36%	75.92%	75.94%	80.82%
NAs	–	–	–	–	87.20%

Source: Our computations based on data from [21–23]. Income and consumption expenditure estimates are grossed up using the surveys sampling weights.

between income and consumption (i.e. consumption propensity) is a relevant economic variable, whose estimate should be among the most significant outcomes of the entire sub-sectoring procedure.

This section shows the results of an exercise aimed at computing consistent indicators for breaking down adjusted disposable income by household typologies and consumption expenditure by consumption purposes and household groups.

In Italy, micro statistics on income and consumption come from distinct sample surveys: the Istat Survey on Households Budgets (Hbs) collects data on households' consumption whereas Istat Italian Statistics on Income and Living Conditions (ItSilc) and the Bank of Italy Survey on Household Income and Wealth (Shiw) focuses on households' income and wealth. Though the Hbs survey focuses mainly on households' expenditure, information on household income and saving is also collected. In fact, households are asked to locate their monthly disposable income and their annual saving within one of several predetermined classes. However, this piece of information has not been released to users as of 2003. Shiw collects also data on households' consumption expenditure but these data are collected only at an aggregated level whereas our objective is to estimate consumption patterns for each household typology.

The matching between surveys is necessary since neither ItSilc (Shiw) nor Hbs contain all the information needed for the analysis. Moreover, the use of ItSilc (Shiw) and Hbs for breaking down separately NAs income and consumption items without any previous matching of microdata, would bring to incoherent results even when survey data are adjusted for differences in time and coverage (e.g. see Table 1). This is probably due to the different sample designs of surveys as well as to the different levels of under-reporting, non-reporting and selection bias.

Table 1 shows the ratio between consumption expenditure and disposable income (i.e. average propensities to consume) for the four Italian macro-regions (North-West, North-East, Centre and South). The first

row presents average propensities to consume computed on Shiw data. The following two rows show average consumption propensities calculated as the ratio between Hbs's consumption and Shiw's (or ItSilc) income. It is worth stressing that surveys' variables were previously recoded in order to align as much as possible to the NAs definitions of income and consumption.

The first point is that surveys data lead to significantly different values of average propensities for Italy. The Shiw's propensity to consume is considerably lower with respect to the figure coming from NAs (around 75% versus 87.2%). However, in line with expectations, Shiw-based propensities decrease moving from the richest (North-East) to the poorest region (South). Results are much more confusing when combining the Hbs data for consumption with the Shiw (or ItSilc) data for income: propensity is largely over one for the South when considering the Hbs-Shiw data source whereas it is even lower with respect to the North according to the Hbs-ItSilc data source.

To improve the coherence between income and consumption expenditure by households' groups, we merge individual survey data before computing breaking-down indicators. The purpose is to obtain a data set (henceforth matched data set) with consistent information on the economic behavior of various groups of households. Matched data set is expected to provide indicators to split up NAs variables by households groups.

Here, statistical matching is applied in order to combine households' income and consumption micro data coming from two independent sample surveys. Alternative sources could be used. For example, records from tax registers could be matched with records from the Hbs sample, through record linkage techniques.

The application pertains Italian data for the 2004, the first year for which ItSilc data were available.

The main steps of the statistical matching procedure consist in: i) choosing the matching variables (the Z variables according to the previous notation); ii) selecting one unit from the Hbs donor sample for each unit of the ItSilc (Shiw) recipient sample; iii) assessing the quality of the statistical matching.

Table 2
Variables selected for the matching of the Shiw (ItSilc) and Hbs datasets – year 2004

Name	Definition	Categories	The variable is used as matching variable	
			Shiw-Hbs	ItSilc-Hbs
Qfood	Quintile of food consumption expenditure	1, 2, 3, 4, 5	✓	
Nmemb	Number of members of the family	1, 2, 3, 4, 5 or more	✓	✓
Njob	Numbers of members with a job	0, 1, 2 or more	✓	✓
Nempl	Number of employees	0, 1, 2 or more	✓	✓
Nedu	Number of members with 11–13 years' schooling	0, 1, 2, 3 or more	✓	✓
Ndegree	Number of members with a university degree	0, 1, 2 or more	✓	✓
Nadul	Number of members aged 40–64	0, 1, 2 or more	✓	✓
Oldfamily	Presence of at least one member aged ≥ 75	Yes/no	✓	✓
Area	Geographical area of resident	Nord-ovest, Nord-east, Centre, South	✓	✓
Rent	The household pays a rent or not	Yes/no	✓	✓

Matching variables are those variables observed both in the Shiw (ItSilc) and Hbs surveys through which the data sets are statistically matched. In order to comply with the Conditional Independence Assumption, it is necessary to select the variables with the strongest explanatory power with respect to both income and consumption [17]. Based on several analysis (multiple regression and Anova models) we selected the matching variables shown in Table 2. It is worth noting that Qfood (household quintile of expenditure on food consumption) – the variable with the strongest explanatory power – is available only for the Shiw-Hbs matching.

Applying the nearest neighbour distance matching, we found one donor unit in Hbs for each unit of (Shiw) ItSilc. The statistical matching procedure generates several matched files, which contain variables from both Shiw (or ItSilc) and Hbs. In order to choose the best-matched file it is customary to compare the distributions of imputed variable (i.e. consumption expenditure) in the donor (Hbs) and in the matched data sets. The matched data set, which shows summary statistics closer to the donor's ones is selected as the best output of statistical matching [17,18]. Table 3 compares the means and standard deviations of Hbs consumption expenditure with the counterpart values computed with Shiw-Hbs or ItSilc-Hbs matched data. The distribution of consumption expenditure seems better preserved in the Shiw-Hbs matched file than in the ItSilc-Hbs one, since the former presents mean and standard deviation values closer to the ones computed on Hbs data. Given that the economic theory predicts a strong correlation between income and consumption expenditure, we also consider high values of the correlation coefficients as an indirect evidence of the “quality” of the data matching [17]. Again, the Shiw-Hbs data set provides the best result, with a value of 0.31 against the 0.23 value obtained on ItSilc-Hbs matched data set.

Table 3

Consumption expenditure: ratios of means and standard deviations computed on matched data sets and on Hbs data (Hbs statistic = 100) – year 2004

	Shiw-Hbs	ItSilc-Hbs
μ	106.95	115.32
σ	101.88	107.14

The lower quality of the integration of ItSilc and Hbs micro data is due to the “weakness” of the matching variables. Households similar with respect to socio-economic characters only (geographical area of residence, number of components, number of components with a degree etc.) do not necessarily record similar values of income or consumption expenditure. The inclusion of income or consumption measures among the matching variables (like consumption or income quintile) would improve the quality of the matching considerably.

Finally, we use Shiw-Hbs matched data to break down NAs income and consumption items by household typologies.

Table 4 shows an outcome of the top-down procedure.

For each Household sub-sector, it is shown the share of income spent on different consumption categories (food, clothing, housing etc.).

Households are classified according to several criteria, namely: the number of children aged less than fifteen years (Nkids) the geographical area of residence (Area), the number of employees (Nemp) the number of retired members (NRetired), the number of females (Nfemale), the number of self-employed (Nself-emp) the number of members aged more than 64 years (Naged), the number of members with a university degree (Ndegree) the household quintile of expenditure on food consumption (Qfood). The last column presents average consumption propensities for the different households groups.

Table 4
Income spent on consumption purposes by Households subsectors, percentage values, Italy 2004

Households subsectors		Food and beverages tobacco	Clothing and footwear	Housing, water, electricity	Furnishing household equipment	Health	Transp. and Comm.	Recreat. culture educ.	Rest. and hotels	Miscell. goods and services	Consump. propensities
Nkids	0	15.1%	6.5%	18.3%	6.1%	2.8%	13.5%	6.4%	6.7%	7.9%	83.2%
	1	17.6%	8.5%	18.1%	10.0%	3.1%	17.0%	8.8%	8.9%	9.4%	101.5%
	2	17.1%	8.7%	18.0%	9.3%	3.7%	14.5%	8.0%	8.1%	8.7%	96.1%
	3 +	21.0%	8.7%	14.7%	6.7%	2.4%	14.2%	7.7%	5.4%	8.9%	89.7%
Area	1	14.4%	6.4%	19.2%	7.3%	3.2%	13.8%	7.3%	7.4%	8.7%	87.6%
	2	12.7%	6.5%	17.3%	6.8%	2.9%	15.3%	6.8%	8.1%	8.0%	84.3%
	3	15.2%	6.3%	18.2%	6.1%	2.2%	12.2%	6.6%	6.9%	7.4%	81.1%
	4	21.4%	9.2%	17.8%	7.7%	3.3%	15.2%	7.1%	5.9%	8.5%	96.0%
Nemp	0	14.7%	5.1%	18.4%	5.0%	2.9%	10.5%	5.1%	4.8%	6.7%	73.1%
	1	16.1%	7.8%	17.7%	7.1%	3.1%	15.9%	7.7%	8.0%	9.0%	92.5%
	2 +	17.1%	9.4%	18.5%	10.5%	2.7%	17.9%	9.2%	9.8%	9.7%	104.7%
Nretired	0	15.2%	8.0%	17.5%	7.9%	2.7%	15.1%	7.7%	8.2%	8.9%	91.2%
	1	16.5%	5.7%	20.0%	5.9%	3.2%	13.0%	5.8%	5.6%	7.3%	83.1%
	2 +	16.6%	5.2%	17.7%	5.1%	3.4%	12.2%	6.0%	5.4%	6.7%	78.3%
Nfemale	0	10.9%	5.0%	18.9%	4.1%	2.0%	11.1%	5.6%	6.3%	6.4%	70.4%
	1	15.8%	6.5%	18.8%	7.4%	2.7%	13.8%	6.2%	6.8%	8.0%	86.0%
	2	16.6%	8.3%	17.5%	6.9%	3.5%	15.9%	8.4%	8.0%	9.0%	94.1%
	3 +	17.3%	8.2%	15.8%	7.9%	3.2%	13.3%	8.0%	6.9%	8.3%	89.0%
Nself-emp	0	17.6%	7.5%	20.4%	7.8%	3.3%	15.2%	7.5%	7.6%	8.7%	95.6%
	1	10.8%	5.8%	12.6%	5.3%	1.8%	11.1%	5.5%	6.0%	6.9%	66.0%
	2 +	8.2%	4.6%	7.8%	2.3%	1.3%	9.7%	4.8%	4.6%	5.3%	48.7%
Naged	0	15.5%	7.9%	17.6%	7.7%	2.8%	15.4%	7.7%	8.2%	8.8%	91.6%
	2	16.2%	5.4%	20.4%	5.3%	3.0%	10.7%	5.8%	4.7%	7.0%	78.6%
	3 +	16.4%	4.4%	18.1%	5.5%	3.4%	12.2%	4.8%	4.8%	6.4%	76.0%
Ndegree	0	17.5%	7.4%	19.3%	7.2%	3.2%	15.3%	7.0%	7.3%	8.6%	92.7%
	1	10.8%	6.5%	15.7%	6.1%	2.2%	11.3%	7.1%	7.3%	7.1%	74.1%
	2 +	8.7%	4.9%	12.7%	7.2%	1.2%	8.8%	6.3%	5.3%	6.4%	61.5%
Qfood	1	10.0%	4.9%	24.4%	3.7%	3.1%	10.4%	4.2%	2.6%	6.7%	70.0%
	2	12.3%	5.6%	21.1%	5.0%	2.5%	11.0%	5.0%	3.8%	6.8%	73.0%
	3	15.9%	6.8%	20.5%	6.8%	2.6%	13.1%	5.7%	6.3%	8.1%	85.9%
	4	17.2%	7.5%	18.6%	8.2%	3.3%	16.1%	7.1%	7.9%	8.4%	94.3%
	5	17.4%	7.9%	14.2%	7.8%	2.8%	15.1%	8.8%	9.3%	9.0%	92.2%
Households		15.7%	7.0%	18.2%	7.0%	2.9%	14.1%	7.0%	7.1%	8.2%	87.2%

It is worth noting that in Italy, neither macro nor micro data sources supply this kind of information. Furthermore, this information is fully integrated in the NAs framework.

The matched dataset provides information for breaking down other relevant NAs items. In addition, the integration of micro data on income and consumption would allow one to perform several households' groupings, thus better meeting the users' various requests.

There is consensus among researchers on income distribution that ItSilc (which is based on a very large sample) provides better estimates of income distribution than Shiw. However, based on available data, it is not possible to get good quality ItSilc-Hbs matched data. The quality of results depends on the quality of both statistical matching and micro-macro reconciliation. To improve the quality of the ItSilc-Hbs match-

ing it would be advisable to include auxiliary information in the matching procedure [17,20]. For example, it would be extremely useful if Istat could enrich ItSilc dataset with data on consumption (e.g. adding few general questions in the survey) and/or Hbs dataset with data on income (e.g. retrieving information from administrative data sources through record linkage). Recently, Istat has been working on the matching of the ItSilc and Hbs micro data, with some preliminary but promising results [24].

5. Developing the bottom-up approach: Towards micro-based Households accounts in Italy

In a full-fledged bottom-up approach, all data sources used to estimate NAs should be made consistent

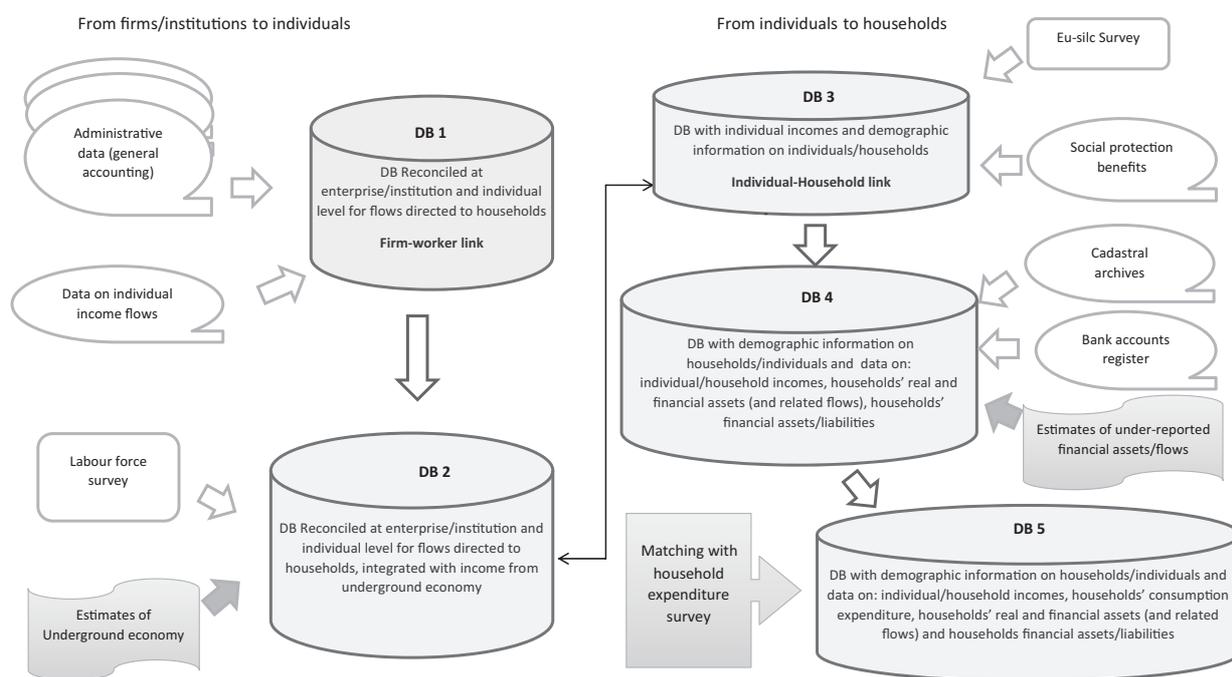


Fig. 1. Framework for the building of micro-macro households' accounts in Italy: from firms/institutions, to households.

at micro level, using matching and calibration techniques [25]. The resulting coherent data sets should be comprehensive for all economic agents, households as well as enterprises or general government entities, and allow tracing all interactions among them. Any imputation, computation and estimation process should be run at the micro level. Balancing methodologies should then be used to achieve consistency at the macro-level [26]. In the bottom-up approach, households' microdata should contribute to define NAs totals.

Though Istat is not planning to develop a full-fledged bottom up approach for estimating NAs, recent methodological developments let envisage a clear advancement in that direction.

Istat has already a consolidated tradition in the integration of several data sources for compiling macro aggregates. The challenge is now to enhance the use of administrative data and deepen the analysis of the firm-worker and individuals-household links. Interlinked records of production units, institutions, earners and households are the basic bricks for developing a full-fledged bottom-up approach.

An ideal framework can be made up by a set of databases, representing different phenomena (income from labour, property incomes and stock, other incomes, consumption), each of them characterized by the possibility of linking each record (and the related flow/stock) to an individual and a household. The pre-

requisite is the existence of business/institution registers and population registers that allow linking individuals to households. Figure 1 shows the different databases and their links.

The left side of the Fig. 1 delineates the steps to be taken for estimating individual earned incomes according to the payer's perspective: the idea is that of estimating the amount paid by each production unit to each worker (be he/she an employees or a self-employed) exploiting administrative and survey data. Administrative archives allow one not only to record per-enterprise/institution values (production, value added), but as well to keep track of all income paid to individuals/households (through compensation of employees, mixed income or property income). These income flows (DB1, Fig. 1) should be estimated starting from enterprises/institution budgets, so the first step deals with the consistency between the income flows directed to individuals/households and the value of flows estimated in the production and generation of income accounts in NAs. In the second step, individual incomes should be integrated with income from hidden economy, whose estimate stems from the linkage and comparison of administrative and survey records. In fact, the analysis of inconsistencies among sources can be used as a signal of hidden economic activity, for example the comparisons of Labour force survey data with social security archives allowed de-

tecting and therefore modelling a significant share of non-registered jobs [27] and related flows. In a similar fashion, under-reported turnover (that enters in production, value added and therefore income earned by households) can be detected by looking at incoherencies among sources [27]. The outcome (DB 2, Fig. 1) would be a database with incomes earned by each employee or self-employed, corrected with imputed income from underground economy. Each record would identify the worker as well as all the production units paying her/his compensations (firm-worker link).

The right side of Figure 1 focuses on Households-Individuals links. Matching administrative and survey data, it is possible to link each individual earning income to the household she/he belongs to, thus identifying individual-household links (DB 3, Fig. 1). Individuals include workers and all individuals receiving other kinds of income like property income (part of it can be detected via fiscal sources) or social security benefits (detected via social security archive). Individuals not receiving any income should be detected as well, since any measure of the distribution of household economic resources must include all people resident in a country, whether or not they are involved in production activity.

A further step would require the link with information on households' real asset (and related flows) and financial assets/liabilities (DB 4, Fig. 1). Cadastral archives could provide significant information: the analysis of real estate at households' disposal could give useful hints to estimate imputed rents as well as registered and non-registered rental flows. Moreover, these data are essential for connecting income and balance sheets data. In recent years, the quality of cadastral data has sensibly improved. Their integration with transaction archives, allowed estimating dwellings and other building average market prices per square meter for detailed categories and areas. These data are used to estimate balance sheets in Italy [28,29]. Moreover, the Observatory of the Real Estate Market (OMI), a Directorate of the Revenue Agency, already linked cadastral archives with individual fiscal declarations for statistical purposes only, suppling some interesting income-wealth integrated analysis [30] Interests are still missing in this set of income variables. Household income surveys can help in estimating financial returns, but these variables (both stocks and flows) suffer heavily of non-reporting and under reporting. It would be ideal to first dispose of financial stocks (from administrative records) and assets and then estimate their returns. Unfortunately, administrative archives on financial assets suffer from non-reporting and under-reporting prob-

lems as well. However, administrative data sources contain more accurate data of financial stocks than surveys do; in Italy, surveys can account for no more than 30% of the total estimated in macro financial accounts [8]. Unfortunately, so far Istat cannot access banking registers, so only estimates through statistical procedures are feasible. The correct micro attribution of financial flows and stocks is one of the challenges in building micro-founded Households accounts. Finally, it is necessary to add data on households' consumption expenditure to give a complete and consistent description of households' economic behavior (DB 5, Fig. 1). In fact, the aim of the overall system should be the estimate of the entire set of Households accounts from generation of income to its use. The study and the evaluation of how consumption (and saving) propensities vary not just in aggregate over time, but by group of households is crucial in evaluating the impact of economic cycle on households and well-being. Unfortunately, in Italy, only the Hbs survey provides data on households' consumption expenditure by purpose. It is therefore compelling to use statistical matching for linking income and consumption micro data (see Section 4).

In this section, we have suggested a framework of interconnected databases, aimed at developing micro-founded NAs. However, it is important to stress that such micro data framework should be used to feed NAs and not for making micro analysis (e.g. to model the behavior of individuals), since the reliability of the integration process is checked at a meso level, not at individual level. Finally, it is worth noting that the development of interlinked micro data systems does not imply the computation of NAs' aggregates as sum of micro data. Some components will necessarily continue to be captured only indirectly through balancing procedures.

5.1. *Present and future plans, in practice*

Although Istat is not planning (at the moment) the building of micro-founded Households accounts in the short term, the recent revision of NAs estimation process as well as the on-going research for the building of integrated micro data sets let envisage a much richer analysis of households' transactions and holdings in the medium term. The key novelty is the more intensive use of administrative archives and their further linkage with records from household surveys. The massive use of administrative data does not imply the uselessness of surveys. On the contrary, surveys allow one to analyze the economic behavior of both individuals and

their household. Furthermore, surveys help detect an important part of underground economy.

Some steps have already been accomplished. In the following, we briefly describe available data sets, which may represent a very useful starting point to build up micro-founded Households accounts.

1. **Frame_Sci.** The Frame_Sci archive records budget data for the entire population of operating enterprises. National accountants harmonize such data with ESA definitions and concepts [7] and correct entries for under-reporting. Frame is the main data source for estimating Corporations' accounts, namely production, intermediate consumption, compensation of employees and remuneration of self-employed. Furthermore, when linked to the individual archives of employees and self employed, Frame allows one to identify all worker-firm links and the firm-worker-household link but only for surveyed households (e.g. in Labour force survey, Eu-silc, etc.). This archive may represent the basement for developing database DB 1.
2. **LFS-Admin** (labour force survey integrated with administrative records). This dataset is built linking survey and administrative records by individuals' fiscal code. The dataset allows national accountants to detect non-registered labour input [26] and could help develop DB 2 together with the Frame_Sci database
3. **SILC-Admin** (Italian survey on income and living conditions integrated with administrative records). Analogously to the LFS-Admin, this dataset allows national accountants to identify registered and non-registered workers. Currently SILC-Admin is used to estimate compensation of irregular employees. This dataset should be included in DB 2 and could represent the bridge to DB 3, since it contains information both on individuals and the household thereof.

In the short run, other works in progress could have a significant return on the accuracy of Households accounts. One of the more significant research project focuses on the matching of micro data on households' income (ItSilc survey) and consumption expenditure (Hbs survey). This matching, already explored in the past [16,17], is now run with the help of auxiliary variables provided by administrative sources [24]. The integration of income and consumption expenditure data is essential to develop DB 5.

Another important research project aims at building an archive for private and public institutions sector,

similar to the one developed for enterprises. This supplementary archive would complete the detection of the labour income flows from the employer side, thus enriching DB 1.

6. Concluding comments

The building of Households sub-sectors accounts is a challenge for national accountants.

The top-down method is based on a strong assumption, namely that NAs provide most reliable estimates than households' budget surveys do. Furthermore, breaking down NAs' totals with patterns derived from several independent data sources may bring to inconsistent results.

Both these limits should be overcome if Households sub-sectors accounts were developed along with NAs, following a bottom-up approach. This solution is quite demanding in that it requires a complete re-thinking of NAs methods.

In order to build sound micro-founded NAs, it is necessary to exploit and integrate all available micro data sources including those relating to households' economic resources. Traditionally, however, data integration has been performed on production processes and on market transaction flows. Conversely, the Household sector has, somehow, been neglected in the integration framework and has often been seen as a residual sector. Therefore, the challenge is to consider Households as the core sector or, at least, to give it the same importance as the other sectors.

In our view, the following steps would help pursue this aim:

- To improve the micro-macro coherence of flows on the observed economy [27] so that, at least for this part of the economy, macro values could be obtained as sum of micro values;
- To improve data collection/estimates on domains that are particularly weak in the micro analysis, especially the flows related to financial assets;
- To build a database where micro data on households' key economic variables are made coherent, using record linkage and statistical matching techniques;
- To introduce a truly independent GDP estimate from the Households side based on households budget microdata and use a SAM [25] instead of a Supply and Use table to balance preliminary estimates.

Istat is investing on the integration of administrative records with the purposes of lowering the burden on respondents (firms and households), reducing costs of data collection and improving the quality of estimates. The process is indirectly fostering the development of micro-founded accounts for Households. In fact, inter-linked records of production units, institutions, earners and households are the basic bricks for building Households sub-sectors accounts via a bottom-up approach.

Acknowledgments

We thank the anonymous reviewers whose comments helped improve and clarify this manuscript.

References

- [1] UNECE. Canberra Group Handbook on Household Income Statistics; Geneva, 2011; Available from: www.unece.org.
- [2] Stiglitz Commission. Report on the measurement of economic performance and social progress; 2009. Available from: www.stiglitz-sen-fitoussi.fr.
- [3] T. Stoker, Empirical approaches to the problem of aggregation over individuals, *Journal of Economic Literature* **31**(4) (1993).
- [4] N.D. Ruggles and R. Ruggles, The Role of Microdata in the National Economic Accounts, *Review of Income and Wealth* **21**(2) (1975).
- [5] United Nations, SNA 1995.
- [6] United Nations, SNA 2010.
- [7] Eurostat, European System of Accounts, ESA 2010.
- [8] M. Fesseau, F. Wolff and M.L. Mattonetti, A cross-country comparison of household income, consumption and wealth between micro sources and national accounts aggregates, OECD Statistics Working Papers, No. 2013/04, OECD Publishing, 2013.
- [9] M. Fesseau and M.L. Mattonetti, Distributional Measures Across Household Groups in a National Accounts Framework: Results from an Experimental Cross-country Exercise on Household Income, Consumption and Saving. OECD Statistics Working Papers, No. 2013/04, OECD Publishing, 2013.
- [10] M. Fesseau and P. Van de Ven, Measuring inequality in income and consumption in a national accounts framework, *Statistics Brief, Oecd* n. 19, 2014.
- [11] H.J. Adler and M. Wolfson, A prototype micro-macro link from the Canadian household sector, *Review of Income and Wealth* **34**(4) (1988).
- [12] M. Fesseau, E. Raynaud, S. Le Laidier and J. Bournay, Building a household-subcategories accounting system using French Micro and Macro statistics, IARIW 30th General Conference, Portoroz, Slovenia, August 24–30, 2008.
- [13] C.P. Mc Cully, Integration of Micro and Macro Data on Consumer Income and Expenditures in: *Measuring Economic Sustainability and Progress: NBER Book Series Studies in Income and Wealth*, D.W. Jorgenson, S. Landefeld and P. Schreyer, eds, 2014.
- [14] N.D. Ruggles and R. Ruggles, A Strategy for merging and matching microdata sets, *Annals of Economic and Social Measurement* **3**(2) (1974).
- [15] I.P. Fellegi and A.B. Sunter, A theory for record linkage, *Journal of the American Statistical Association* **64** (1969), 1183–1210.
- [16] A. Coli, S. Colombini, M. Di Zio, M. D’Orazio, I. Faiella, I. Siciliani, G. Sacco, M. Scanu and F. Tartamella, 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, 2006.
- [17] M. D’Orazio, M. Di Zio and M. Scanu, *Statistical matching: Theory and Practice*. Wiley & Sons. New York, 2006.
- [18] S. Rässler, *Statistical matching: a frequentist theory, practical applications and alternative Bayesian approaches*, Lecture notes in statistics, 168 – New York Springer, 2002.
- [19] W.L. Rodger, An evaluation of statistical matching, *Journal of Business and Economic Statistics* **2**(1) (1984).
- [20] A.C. Singh, H. Mantel, M. Kinack and G. Rowe, Statistical matching: Use of auxiliary information as an alternative to the conditional independence assumption, *Survey Methodology* **19** (1993), 59–79.
- [21] Bank of Italy (2006) Italian Households budgets in the year 2004, Supplementi al Bollettino Statistico, Indagini campionarie, Anno XVI Numero 7–17 Gennaio 2006.
- [22] Istat. Households consumption expenditures, year 2004, *Annuario* n. 11, 2006.
- [23] Istat. Cross sectional UDB (User Data Base) ITSILC 2005 - Version: 2005-1 del 22-11-2007.
- [24] G. Donatiello, M. D’Orazio, D. Frattarola, A. Rizzi, M. Scanu and M. Spaziani, The role of the auxiliary information in Statistical Matching Income and Consumption, *New Techniques and Technologies for Statistics (NTTS)*, Bruxelles; 02/2015, 2015.
- [25] Eurostat, Handbook on social accounting matrices and labour accounts, Eurostat Working paper, Population and social conditions 3/2003/E/N°23, 2003.
- [26] R. Stone, D.C. Champerowne and J. Meade, The precision of national income estimates, *Review of Economic Studies* **9**(2) (1942).
- [27] C. De Gregorio and A. Giordano, The heterogeneity of irregular employment in Italy: some evidence from the Labour force survey integrated with administrative data. Istat working papers n.1/2015.
- [28] Eurostat-OECD, *Compilation guide on land estimation*, 2015.
- [29] Istat, *The stock of non-financial assets by institutional sector*, 2015.
- [30] Agenzia delle entrate. Gli immobili in Italia (yearly releases), http://www.agenziaentrate.gov.it/wps/content/Nsilib/Nsi/Agenzia/Agenzia+comunica/Prodotti+editoriali/Pubblicazioni+cartografia_catasto_mercato_immobiliare/Immobili+in+Italia/.