The Role of Micro Data in National Accounts
Towards Micro-founded Accounts for the Household Sector

Alessandra Coli (University of Pisa)
Francesca Tartamella (Italian National Institute of Statistics)

The Role of Micro Data in National Accounts
Towards micro-founded accounts for the household sector

Alessandra Coli[1], Francesca Tartamella[2]

Draft - This version: March 2015

Abstract

This paper deals with the issue of micro and macro data integration when the purpose is to build micro-founded accounts for the household sector. The last decade debate on progress and well-being has stressed the necessity of putting people first, also within the National Accounts (NAs) system. This means giving more attention to the households sector and sub-sectoring seems to be the straightforward solution. However, whatever the chosen sub-sectoring, it would hardly satisfy users' requests, since these change across users and over time. An alternative solution could be the adoption of a basic accounting system for households with satellite tables showing the accounting of a variety of sub-sectors. However, this solution is practicable only if households' macro aggregates are micro-founded. Thus, the integration of households' micro data in the NAs' macro aggregates becomes crucial. The challenge is to find the best possible way to reconcile or align them.

1. Introduction

This paper focusses on the integration of micro data in NAs, with the purpose of developing micro-founded accounts for the households sector.

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. National accountants have tried answering 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 the distributive processes or the analyses by industry and products). On the other hand, they have disentangled transactions by finer homogeneous groups of *transactors*, developing sector and sub-sectors accounts. The result of this double-sectoring process is a very rich source of information, which, however, suffers from two major limits.

The first limit is that households (i.e. people) play a secondary role within the NAs framework. Actually, NAs focus mainly on domestic production processes and on the market transaction flows. Supply and use tables give a detailed picture of production costs and income generation by industry and show in detail the composition of the supply and uses of goods and services. These tables are extremely useful for the economic analysis. For example, input-output tables are integrated in macro-economic models for assessing the direct and indirect effects

---

of changes in prices or tax rates on the values of supply or use. At the opposite, we know very little on households' behaviour. Households is a sort of "residual" sector, still far from being homogeneous. The sector encompasses a wider range of actors than biological persons; it includes non-profit institutions, unincorporated businesses and personal associations. Finally yet importantly, NAs limit to describe the economic behaviour of a representative household, without any further insight into the heterogeneity of the sector. NAs do not examine the economic behaviour of significant and critical population subgroups such as elderly, families with children, poor households etc. Actually, it is striking the imbalance between the richness of industry-based analyses and the absence of any insight in households' economic behaviour. The secondary role of households in NAs is mirrored in the statistical resources that are used. More often, national accountants use households' data less intensively than business or government data for the estimate of GDP. Furthermore, in contradistinction to most of other sectors, not all constituent micro data of the households sector are derived from family or individual files or accounts. Many of the data come from the records of business or governments. Overcoming this first limit implies shifting in focus from firms to households, first enhancing the use of households' data in the estimation process of NAs.

The second limit is the users' difficulty to find and fully understand the content of sectors' key variables. This is due to the presence of imputations, attributions and accounting rules typical of national accounting, which result hardly comprehensible to users and not sufficiently explained. In addition, more often, household' accounting data are disseminated in a way that does not facilitate users. Whereas it is immediate to find statistics on production, resources and uses by industry, much harder it is to find data on households' disposable income, consumption or saving. OECD section on "National Accounts at a glance" is a valuable exception (www.oecd.org).

The last decade debate on progress and well-being has stressed the necessity of putting people first, also within the NAs system. This means giving more attention to the households sector and again sub-sectoring seems to be the solution. However, whatever the proposed sub-sectoring, it would hardly satisfy the users' requests, since the kind of information-required changes across users and over time. Conversely, there is the risk that adding too many details within an already complex system will ultimately make NAs even more inaccessible to most users.

In our view, another viable solution could be the adoption of a basic accounting system for households with satellite tables showing the accounting of specific groups of households. This solution would be practicable only if the estimates of households macro accounts are micro-founded and if the relationship among macro and micro is made clear to the users. Thus, the integration of households' micro data in the NAs' macro aggregates becomes crucial. The challenge is to find the best possible way to reconcile or align them.

The paper is structured as follows. Section 2 provides a quick overview on households accounting within NAs. Section 3 focus on households sub-sector accounts and the building of a micro data set where data on households' budgets from different sources are integrated and made coherent. Section 4 focuses on the role of microdata in the Italian NAs pointing out weaknesses and strongpoints for the building micro-founded households' accounts. Section 5 contains some concluding remarks.

2. Households accounting in official statistics: the micro and the macro perspective

In official statistics, two different kinds of data sources provide information on household income, consumption and wealth. On the one hand, micro sources (surveys, administrative records, censuses) supply information at the level of specific individuals and households. On the other hand, NAs provide measures for the country as a whole and for groupings of transactors (sectors), there including the households sector. Both the approaches have limits and strongpoints.

Micro data allow 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 (UNECE, 2011) represents the international guide for producers and users of micro data on household income. Nevertheless, micro sources typically fail to record all types of income flows: 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) and they may fail to capture all income from shadow economy (although they provide valuable information to detect non regular labour units). These are important limits, especially when the purpose is that of comparing economic well-being of people living in 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 source.

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. Finally, the 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. One strongpoint 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. Furthermore, it is possible to analyse the households' economic behavior vis à vis the other sectors, given the consistency and coherence of the overall system. The SNA offers 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 (Ruggles R. and N. Ruggles, 1975). In fact, supply side flows have to be consistent with flows from the demand and the income side, moreover each flow has to be recorded in double accounting, so that incoming flows must always counterbalance outgoing flows. The estimate of aggregates and flows stem from the comparison of different data sources and this method allows to point out each data source's deficiencies and to identify the missing pieces of the puzzle to build a complete image of the economic 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 analyse the economic behaviour of a representative household (not considering the problem of aggregation over individuals, Stoker 1993). This depends on the fact that official statistics institutions never put households' accounting among priorities. Also nowadays, the transmission program[3] 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 (for example with respect to specific distributive flows originating from the General Government).

In short, macro data assure comprehensiveness, coherence and international comparability but cannot provide information on the distribution of economic resources among people; while micro data allow calculating distribution indicators but do not cover all people's economic resources and suffer of a certain degree of inner inconsistency.

Over the years, the macro and micro approaches have developed separately, sometimes leading to divergent results. In order to bridge macro and micro perspectives, 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 indexes 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 income from hidden economy, benefits in kind received by governments or other imputed transactions included in the macro aggregate. Ideally, NAs measures should incorporate distribution information on households' accounting. Performing distributive analyses within the NAs framework (instead of using micro data) would allow 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.

In order to incorporate distribution information in NAs, ESA 95 suggested partitioning the Household sector into sub-sectors according to the family's largest income category (ESA 95 §2.79, p. 29). Furthermore it was encouraged the development of Social Accounting Matrices (ESA95, § 8.133-8.155) as the proper accounting framework to better focus on the households' economic behaviour. The Leadership Group on Social Accounting Matrices and Labour Accounts significantly contributed to this kind of analysis. The main outcome was a Handbook with the guidelines for the building of social accounting matrices and the estimate of pilot SAMs for nine European countries (Eurostat 2003).

In 2009, the Stiglitz-Sen-Fitoussi Commission on the measurement of economic performance and social progress (Stiglitz Commission, 2009) renewed attention on the topic.

The Commission recommends to give prominence to the distribution of income, consumption and wealth (SSF report, Recommendation 4, p. 40), adding how "ideally distributional measures should be compatible in scope with average measures from the national accounts" (SSF report, § 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 (SSF report, § 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) estimates with the ultimate purpose of estimating some significant NAs items by

---

[3] 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.

households sub-sectors. Fesseau M. et al. 2013a and 2013b present the main results of the Expert Group whereas Mary Fesseau and Peter Van de Ven (2014) present some inequality measure on the distribution of income based on such results.

3. Building of micro-founded households accounts

Building households' sub-sectors or social accounting matrices implies facing the problem of how integrating micro information from households data sources without losing the consistency of the overall accounting system. Two main approaches can be followed:

i. Top-down approach: NAs items are considered as the standard to be broken down using information from individual-based data (Fessau *et al.* 2013-b). This approach is based on the assumption that NAs provide most reliable estimates with respect to 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 as an external yardstick.

ii. Bottom-up approach: micro data on households' budgets are directly used to estimate NAs. The underlying assumption is that the divergences between existing micro and macro data must be bridged meeting somewhere in the middle (Adler H. J and M. Wolfson 1988). 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 (Ruggles R. and N. Ruggles 1975).

We can distinguish two common steps in the practical translation of both the approaches.

First, it is necessary to reconcile existing macro and micro statistics on households budgets, given the significant divergences that often emerge between NAs totals and totals coming from sample surveys micro data sources. In principle, data in micro data sources should be consistent with the macro data of NAs. However, this does not happens for several reasons such as the different time periods, different coverage (populations) or a different content of the compared variables. Unfortunately, even where the sectoring of the national accounts corresponds precisely to the universe represented by the micro data source, differences may continue to be significant (e.g. see Coli, Tartamella 2008, Fesseau *et al.* 2008).

Second, it is necessary to integrate data from different data sources to build a consistent database of individual data on households' budget variables (henceforth *matched-dataset*). Unfortunately, no single micro data set contains all the information required for estimating all NAs variables by households' sub-groups (Coli A. and F. Tartamella 2008, Fesseau *et al.* 2008, Mc Cully 2014), so that it is necessary to use several data sources. However, different micro data sources often provide conflicting estimates on a same phenomenon or inconsistent estimates on inter-connected phenomena (e.g. income and consumption expenditure). National accountans know very well the difficulties connected with the merging of data coming from different data-sources; indeed, one of main national accountans function is to find the best statistical method to resolve such conflicts and inconsistencies within the overall framework of NAs.

The Top-down approach continues with the breakdown of the households' NAs items according to the patterns stemming from the matched dataset. The Bottom-up approach requires establishing a method for including matched data in the NAs estimation process. As a first step, it would be useful to enhance the use of these data for the estimates of all NAs items where households are involved. In a long-term strategy, data from the matched-dataset could be merged and integrated in the NAs framework during the balancing step (which should be run in a SAM instead of in a Supply and Use tables).

Whatever the chosen approach, the construction of the matched dataset would give the possibility of estimating at least the NAs key variables for households grouped according to different criteria.

3.1 Construction of the matched data set

Ideally, one would like to combine economic, social and demographic information from a wide variety of different sources for every household and even for individuals within the household. Assuming not to have confidentiality problems, we could for example imagine of using record linkage techniques[4] to retrieve information from census records, taxes records or social security records. In fact, record linkage techniques aim at identifying pairs of records, in different data sources, which refer to the same entity (person, household, enterprise, etc.). 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

---

4

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 complications may occur: 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 which agree may refer to different people. Formalizing the linking procedure into a statistical model, it is possible to evaluate the matching by measuring the probability of generating false-matched-pairs and false-unmatched pairs (Fellegi and Sunter 1969). In deterministic record linkage, the decision whether a pair of records refer to the same entity or not is based on *ad hoc* rules whereas in probabilistic record linkage the decision is based on statistical models.

Unfortunately, record linkage can be used only when compared data sources are supposed to contain the same entity i.e. when the linkage involve one sample survey and one administrative data source or two administrative data sources covering the same population. For example, we could use record linkage techniques to combine information from a sample survey on households' consumption expenditure with data on income coming from tax records. At the opposite, 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 consumption expenditure information from two different sample surveys (Coli Tartamella 2008). In this context, the use of statistical matching seems appropriate. Statistical matching is used to link independent samples of data *A* and *B*, by means of some variables common to both data files (D'Orazio et al. 2006). 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:

i)   *A* and *B* contain both a set of common variables Z and a set of specific variables Y and X respectively;

ii)   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 (Rässler 2002).

This procedure can be correctly applied only when the Conditional Independence Assumption (CIA) between Y and X given Z holds. In fact, the matching procedure generates conditional independence between the variables not jointly observed even when such variables are conditionally dependent in reality (see especially D'Orazio *et al.*, 2006; Rässler 2002; Rodgers 1984). 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 Rässler, 2002 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 Ruggles, 1974). 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 D'Orazio *et al* (2006); Singh *et al* (1993).

Imputation by regression is less satisfactory than statistical matching, when the purpose is transferring complex sets of information from one source to another. Ruggles R. and N. Ruggles (1974b) recalls and clearly describe the advantages of the use of statistical matching with respect to imputation by regression.

The building of the *matched-dataset* requires the use of both record linkage and statistical matching techniques.

Box 1 Matching households data on income and expenditure from sample surveys – an application for Italy

We apply statistical matching in order to merge the micro data on households' income (EUSILC or SHIW) with the micro data concerning households' consumption (HBS). Our purpose is to obtain a data set with consistent information on the economic behaviour of groups of households, including the analysis of consumption patterns by households' subgroup.

The matching between the two surveys is necessary since neither EUSILC (or SHIW) nor HBS contain all the information needed for the analysis, either for income or for consumption. Actually, SHIW collects also data on households' consumption expenditure but only at an aggregated level.

Furthermore, the separate use of EUSILC (SHIW) and HBS to breakdown income and consumption expenditure may bring to incoherent results. Table 1 provides an example of such incoherence comparing the consumption propensities calculated on a single data source (SHIW) and consumption propensities obtained by using EUSILC (SHIW) and HBS without any previous matching process

Table 1 Average consumption propensities by geographical Area - Italy, 2004.

Data source		North-West	North-East	Centre	South	Italy
Shiw	weighted values[a]	72.93%	71.54%	75.52%	81.08%	75.09%
Hbs-Shiw	weighted values[a]	100.37%	87.76%	82.06%	107.11%	94.79%
Hbs-Itsilc	weighted values[a]	86.31%	84.36%	75.92%	75.94%	80.82%

Own elaboration based on data from Bank of Italy and Istat.

[a] Income and consumption expenditure estimates are grossed up using the surveys sampling weights

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 EUSILC (SHIW) recipient sample; iii) assessing the quality of the statistical matching.

Within common variables, those more strictly connected with income and consumption are selected in order to fulfil the CIA. Applying the nearest neighbour distance matching, we found one donor unit in HBS for each unit of EUSILC (SHIW). The statistical matching procedure generates matched files, which have the same dimension as the recipient dataset (EUSILC or SHIW). In order to choose the best matched file it is 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 was to impute HBS consumption expenditure, we compared summary statistics on total consumption expenditure calculated with HBS data and with each matched file data (see Coli 2009 for details). Distributions are well preserved in the SHIW- HBS matching. The EUSILC-HBS matching gives worse results. This is due to the "weakness" of the matching variables in explaining household income and consumption. To improve the quality of the matching we should include auxiliary information in the matching procedure. For example, it could be extremely useful if Istat could enrich EUSILC with data on consumption (adding few general questions in the survey) and HBS with data on income (e.g. retrieving information from administrative data sources through record linkage).

Finally, using matched data we could calculate consumption propensities by households groups and consumption purposes. Table 2 shows some results.

Table 2 Income spent for consumption purposes by Households subsectors, percentage values, Italy 2004.

		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
Nkids	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
Nemp	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
Nretired	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
Nfemale	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
Nself-emp	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
Naged	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
Ndegree	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

4. The role of micro data in the Italian National Accounts: towards micro-founded households accounts.

The Italian statistical institute, has already a consolidated tradition in the use of administrative sources from the enterprise side and the National Accountant usually consistently integrate several sources in compiling macro aggregates. The challenge is now to enhance their use, exploiting the information to deepen the analysis of the link enterprise-household so that they can be the first step toward the building of micro-funded households accounts.

Taking into accounts the current statistical context and especially some recent development of data production, and in the availability of data (administrative or from surveys), it is possible to envisage the step to be done to build micro founded household accounts in Italy (see figure 1).

The first domain of integration of administrative data on households can concern the labour market, the objective being to detect all labour input and related income flows between productive system and households. Administrative sources have to be harmonized with NA definitions and have to be consistent with outlays recorded by enterprises as compensation of employees. It happens, in fact, that commercial accounting have to comply with different regulations in terms of social security obligations, that force them to keep individual records of each person employed in the production, and general accounting, that lead to profit and loss accounts that record total labour costs. Now it is the latter that nourish Structural Business Statistics variables[5], the base for NA compensation of employees estimates. But social security or fiscal data, available at individual level, does not have necessarily the same income concept of general accounting, since they have to answer to different legislation (they follow mainly a cash concept and not all income are imposable to taxes or contributions). So individual data have to be reconciled with total firm compensation of employees to obtain individual flows of income. Company accounts also lead to estimate compensation of self employed. Also these flows can be traced down to individual level.

Moreover, administrative data can account about registered employment and income only. To detect also non registered labour input (or at least the great part of it) and non registered income flows the integration with household survey, like labour force survey, can be crucial. With these information it is possible to have an exhaustive picture of the labour market, so estimating labour input and related income flows.

The following domain of integration is fiscal sources (with data on dividends and registered rents) and archives of social security benefits. Moreover, the analysis of dwellings at households disposal can give useful hints to estimate non registered rental flows. These archives can complement the set of income variables, when harmonized with NA definitions and concepts.

Interests are the main flows that remain missing in this set of income variables. Household income survey can help in estimating financial returns, but also in surveys these variables (both stocks and flows) suffer heavily of non reporting and under reporting. Moreover, financial stocks are quite concentrated, so even when households correctly answer to surveys it is not necessarily easy to find reliable ways to impute at a micro level the corresponding flows. It would be ideal to first compute financial stocks and assets and then estimate their returns. Unfortunately banking registers are not accessible to the statistical institute, so only estimates trough statistical procedures are feasible. The link with real asset would be, instead, straightforward, since real assets are recorded in administrative archives. Completing the framework of income flows to balance sheets would allow to complete the domain of analysis to use of income, investment and savings and therefore to balance sheets. The correct micro attribution of financial flows and stocks is one of the challenges in building micro-funded households accounts.

To complete the information system on households also consumption data have to be integrated: the aim of the overall system should be to allow to estimate the entire set of accounts from generation of income to its use. The study and the evaluation of how consumption (and saving) propensities varies not just in aggregate over time, but by group of households is crucial in evaluating the impact of economic cycle on households accounts and well-being. Unfortunately, there is not profusion of detailed (on consumption functions) individual data on household expenditure: so the only information source is the household expenditure survey. It is therefore crucial to use statistical matching between the income information system and the national expenditure survey. Also this survey should be adjusted, both in concepts (to be consistent with NA definitions) and in amounts, since consumption totals for some items fall far from aggregated macro value, and any administrative data on some

---

[5] Both in case that SBS is estimated using administrative records and in case that direct survey on firms are performed, since enterprises tend to answer to surveys using general accounting data.

consumption expenditure, where existing, should be integrated (for example, there can be records on purchase of some consumption durables). Moreover, the micro (administrative) information on income can be a helpful auxiliary variable in the statistical matching between the income and the expenditure source.
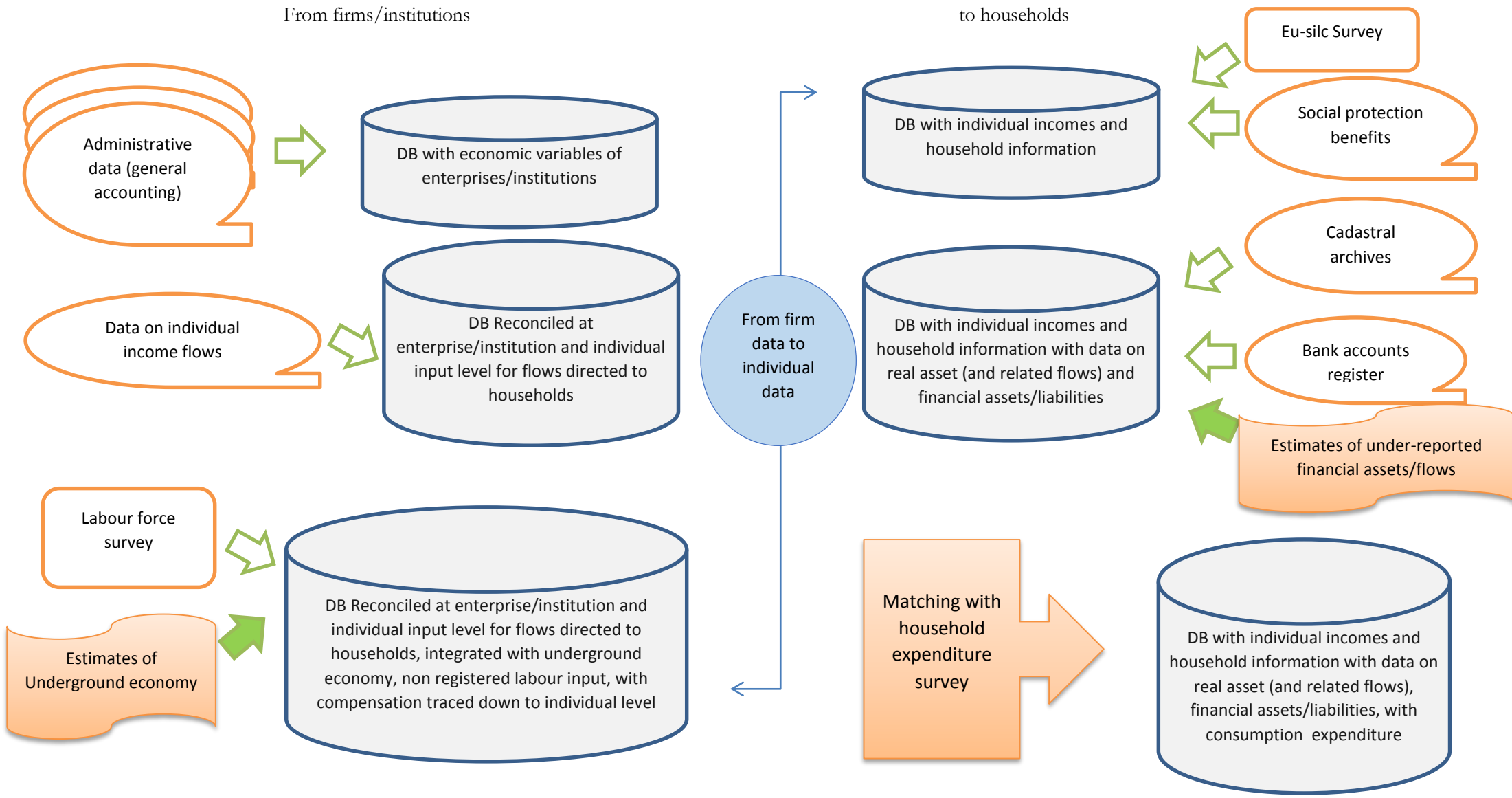
Of course there are variables that are linked to NA concepts and have to be imputed (i.e. Fisim, property income distributed to insurance policy holders, etc), but the challenge is to limit imputation as much as possible, not to affect the general income distribution and try to impute them at least with an acceptable meso-level reliability.

The intensive use of administrative data allows to build useful database on firms and households, trough micro linkage. Building databases with administrative records on individuals and households has the disadvantage of handling heavy database, and their management can be quite complex, since it necessarily includes a huge amount of records. Moreover, it is essential to monitor the legislation on which the administrative data are based, which, too, can be quite complex and requires fiscal, and accounting competences and not just statistical ones. But, on the other side, building such databases lowers the burden on firms and households, since only information that can not be elsewhere recovered have to be asked.

The construction of a database with the characteristics and involving all steps and domains of integration envisaged above is quite demanding. The feasible compromise could be to build a database from surveys: income survey and labour force surveys could be used together, to expand the number of households (of course re-computing sample weights), integrated with administrative data as much as possible and made coherent with data from counterpart sectors, imputing missing and non registered components and domains out of scope of household survey (persons not living in private dwellings or illegal immigrants). Such database, harmonized with NA concepts, can be the ideal instrument to disaggregate NA values according to any desired households characteristic (quintile, household composition, age etc).

Nevertheless it is important to underline that according on how far the imputation procedure is pushed, such a database should not be used for micro analysis (i.e. to model the behavior of individuals), but for macro and meso analysis.

.

Figure 1: Framework of micro-macro households accounts in Italy: from firms/institutions, to households

From firms/institutions

to households

Administrative data (general accounting) → DB with economic variables of enterprises/institutions

Data on individual income flows → DB Reconciled at enterprise/institution and individual input level for flows directed to households

From firm data to individual data

Labour force survey →

Estimates of Underground economy → DB Reconciled at enterprise/institution and individual input level for flows directed to households, integrated with underground economy, non registered labour input, with compensation traced down to individual level

Eu-silc Survey

Social protection benefits → DB with individual incomes and household information

Cadastral archives

Bank accounts register → DB with individual incomes and household information with data on real asset (and related flows) and financial assets/liabilities

Estimates of under-reported financial assets/flows

Matching with household expenditure survey → DB with individual incomes and household information with data on real asset (and related flows), financial assets/liabilities, with consumption expenditure

4.1 Present and future plans, in practice

Although Istat is not planning (at the moment) the building of micro-funded households accounts, 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 in the medium term. The key is the more intensive use of many administrative archives, integrated in a statistical process, not just among them, but also with household surveys.

Some steps have already been developed in Istat in the construction of micro databases that can be a very useful starting point to build up a micro founded analysis on households (even when the data are not conceived to focus on the household sector, the richness in details allows to analyse flows directed to households keeping information on households characteristics) .

1.   Frame-SBS: Istat built an archive of small-medium (1-100 workers) enterprises with economic variables derived from administrative data[6].NA adjust these data to take into account under-reporting of entries[7] at a micro level, though reliable at a meso level, harmonize them with ESA definition and use them as the main source for NA estimates on the enterprise sector: production, intermediate consumption, compensation of employees and remuneration of self employed. Even though this archive has not been built to analyse the household sector, it is possible to use it for this purpose, identifying each persons employed in each firm (as employee and self-employed), and following the income flows from the firm to the specific household[8]

At the moment there is still not an archive of household composition (i.e. the link from the individual to each household), so it is possible to analyse these flow directed to households by the characteristics of the individual, not the characteristics of the household. But it is possible to use an household survey as a representative share of the population, and therefore direct these income flows to each individual of the survey. The resulting income then can be studied according to household composition.

2.   Integrated sample Lfs-Admin (labour force survey statistically integrated with administrative records). This sample allows to detect non registered labour input within a statistical integration process that corrects employment level bias: possible under-coverage of employment in Lfs, and possible administrative over coverage of registered jobs and lack of coverage of non-- registered jobs (see De Gregorio, Giordano 2015, AA.VV. 2015). This integrated sample entered in the NA employment estimation process. It is an important source to detect the link firm (or institution)-household for sample registered employed. For example, it is possible to link the data derived from business data (point 1 above) with this integrated sample, obtaining a full picture of flow that households derive from the labour market, in a registered and non registered way.

3.   Integrated sample Silc-Admin (Italian Silc integrated with administrative records). Analogously to the Lfs-Admin integrated sample, this sample allows to identify registered and non registered workers and was used to estimate a wage differential (by Nace) between registered and non registered employees that entered in NA estimates compensation of irregular employees (see IStat 2014).

Istat is continuing investing on data integration and use of administrative records, both with the purpose of lowering the burden on respondent (firms and households) and costs of data collection, and improving the quality of statistical output. This has an impact also in the domain of household statistics.

There are several work in progress, already at an advanced state, that in the short run can have a significant return on the improvement of micro-founded household statistics.

a)   Improvement of Silc-Admin. Extension to this integrated survey of the probabilistic model to determine the correct status in employment and eventually impute some income flows to (non registered) positions resulting from the model. In this way each individual in the It-Silc sample has a checked status in employment also in terms of industry (derived from administrative source) and the so estimated number of workers are fully

---

[6] The Italian productive sector is characterized by a very high number of small enterprises. This fragmentation has always been a big limitation in data collection. Now an exhaustive database on all enterprise (in Nace B-S) based on administrative data (accounting and fiscal data) has been built (more than 4,5 million enterprises) on the main economic variables in the SBS domain. A specific attention has been paid to compensation of employees estimates. Using social security archive as auxiliary sources it was possible to adjust compensation of employees to exclude some mis-classification of items that, in NA and SBS definitions, are more properly classified in intermediate consumption. The most relevant (in terms of amount involved) correction concerned the registration of outworkers and agency workers that some enterprises classified among compensation of employees, instead of intermediate consumption (see Arnaldi and al, 2015)

[7] This component of underground economy accounted for about 5.7% of GDP in 2011.

[8] Keeping in mind that this operation involves some degree of approximation: for compensation of employees, since individual data have to be reconciled with aggregated firm data, for self employed since the imputation of non-reported entries is reliable at meso level.

comparable with NA employment. Compensation of (registered) employees flows have been reconciled with paying institution data. Also for self employed, income flows of registered workers have been reconciled with firm profits, also for the component deriving from under-reporting of firm entries.

Silc-Admin, nevertheless, remains inadequate on non registered flows from real assets, (non registered rentals of buildings, mainly dwellings); in fact NA estimate a relevant share of non registered rents from dwellings, only part of it is recorded by the survey. Only the integration with cadastral archive could allow to estimate non registered flows..

Another crucial step of integration is about property income from financial assets, which is substantially underestimated with respect to NA totals, which are derived from financial accounts (estimated by the Central Bank), supervisory agencies of financial intermediaries[9], public administration, the balance of payments: these data (almost[10]) exhaust the total interest flows paid in the economic system. In the short run the chance of obtaining bank account registers are weak, so the only attempt can be the estimate of under reporting and under-reporting trough a statistical method, looking for auxiliary information.

b)   Integration with expenditure data: Istat had already experienced in the past (see Coli Tartamella (2008), Coli et Al (2006))) and is improving now different procedures to perform a statistical matching between the Italian Silc and the expenditure survey. The procedure that uses only survey data has been compared with a different procedure that stratifies donors and recipient households on the base of income from administrative (fiscal) sources (on mainly labour compensation and social benefits), i.e. with the underlying hypothesis of uniform (or non influent) distribution of non registered and financial flows. The results have still to be evaluated, both in terms of relation with income and detailed consumption items and especially (in our view) in terms of coverage of NA values.

c)   Frame-S13: Istat is experimenting the production of an archive with the main economic data on all institutional units of S13 (public administration) sector, similarly of what has been done for the archive of small-medium enterprises. European regulation requires very detailed information to produce Public Finance statistics. The accounting system of institution part of the public sector is fragmented and data collection on this sector is very labour intensive. The organization of a coherent framework of the main economic variables could be a helpful support in production of NA on S13. This supplementary archive, can complete the picture of the labour market income flows, complementing the information from the employer side, than then can be attributed to the households.

In the long run Istat will develop the archive Archimede[11] (integrated archive of economic and demographic micro data), that should integrate (in a coherent framework) the complete set of administrative sources for each resident individuals, both fiscal and from social security. This archive is supplemented with demographic information (on the individual and his/her household). The target income variables are not necessarily the one consistent with NA definitions, but the archive should contains all information necessary to compute (or estimate) the necessary adjustment. The archive will include all information on registered income flows (remuneration of labour input, social benefits, transfers, real asset income flows). It may remain inadequate on interest flows, that do not have to be included in fiscal declaration, but there is the chance that Istat can obtain a sample of bank account registry.

This archive will be integrated with the information from the household expenditure survey, so that it will be possible to have a complete record with income and consumption. The micro linkage can be performed only for the household surveyed –(and the results expanded to population with sample weights), but the target is to expand the domain of analysis and find a way to expand the sample trough statistical methods (if not estimating consumption expenditure for all population, at least for some items).

This archive aims to be the base not only for NA estimates, but also to integrate Silc estimates, as a base for microsimulation models and all analysis and prevision tools based on households.

Of course, for NA estimates, this archive have to be integrated with the non registered economy components and be reconciled with flows estimated for counterpart sector.

What is important to underline is that the massive use of administrative data does not imply that it is possible to do without household survey: the analysis so far proved that household survey help in

---

[9] The current legislation imposes supervision of all Italian banking and insurance enterprises. The information collected for this purpose by the supervisory authorities provides an exhaustive information base for the financial sector property income flows.

[10] The only interest flows that are left out are those that do not have as counterpart sector financial corporation or the public administration or the rest of the world, i.e. interests on commercial credits/dets. Also stock and bonds issued by corporation are monitored.

[11] See Garofalo 2014

detecting an important part of underground economy (and help in estimating the remaining!), this hold for Lfs and Silc at the same way

5. Concluding comments

In order to build sound micro-founded NAs, it is necessary exploit and integrate all available micro data sources including those relating to households' budgets. Besides, NAs has always been the ideal framework for integration of sources, having the purpose of an exhaustive comprehensive and consistent illustration of all the economic system.

Traditionally, however, data integration has been performed on production processes and on the market transaction flows. So the household sector has, some how, been neglected in the integration framework and often seen as a residual sector.

Now it is clear that analyzing the household sector does not only gives useful insights on the distribution and use of what is produced, but also on the way income distribution takes place (i.e. evenly or not) and may in turn, affect the other agents' behaviour. So the challenge is to consider household as the core sector or, at least, to give it the same importance as the other sectors.

Several steps have to be undertaken to pursue this aim:

a) to improve as much as possible the micro-macro coherence of flows on the observed economy so that the macro value can be obtained as sum of micro values;

b) to improve data collection/estimates on domains that are particularly week in the micro analysis, especially the flows related to financial assets;

c) to build a database where micro data on households' key economic variables are made coherent, using record linkage and statistical matching techniques to merge data from different sources;

d) to introduce a truly independent GDP estimate from the household side and use a SAM instead of a Supply and Use table to balance preliminary estimates.

The national statistical institutes has to decide to which degree pushing this work that involves an intense investment, but can improve the general statistical framework and enhance not just the analysis of households, but the quality of NAs as a whole, since integration has always some feedback on the quality of the data.

Bibliography

Adler H. J. And M. Wolfson A prototype micro-macro link fro the Canadian household sector, Review of Income and wealth, 34(4)

Arnaldi S., Baldi C., Filippello R., Mastrantonio L., Pacini S., Sassaroli P., Tartamella F. (forthcoming) *The labour cost variables in the building of the SBS Frame,* Rivista di statistica ufficiale, Istat.

AA.VV. *Soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione.* Final report of Istat Working group on the Development of new methods for employment estimates for National accounts benchmark. Mimeo, forthcoming as Istat working paper. Roma:

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.

Coli A., Tartamella F. (2008): Income and consumption expenditure by households groups in National accounts, Paper presented at the IARIW Conference, Portoroz, Slovenia, August 2008.

Coli A. (2009) Combining household data on income and expenditure from sample surveys and National accounts, NTTS 2009.

De Gregorio C., Giordano A., 2015, *The heterogeneity of irregular employment in Italy: some evidence from the Labour force survey integrated with administrative data.* Istat working papers n.1/2015

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

Eurostat (1995) European System of Accounts

Eurostat (2003) Handbook on social accounting matrices and labour accounts, Eurostat Working paper , Population and social conditions 3/2003/E/N°23.

Eurostat 2013, European System of accounts 2010

Eurostat 2014 European system of accounts - ESA 2010 - Transmission programme of data

Fellegi I.P. and Sunter A.B. (1969) "A theory for record linkage", *Journal of the American Statistical Association*, 64, pp. 1183–1210.

Fesseau, M., E. Raynaud, S. Le Laidier, J. Bournay (2008), *Building a "household-subcategories accounting system" using French Micro and Macro statistics*, IARIW 30th General Conference, Portoroz, Slovenia, August 24-30.

Fesseau, M, Wolff F. and Mattonetti, M-L (2013a), 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.

Fesseau, M. and M. L. Mattonetti (2013b), 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.

Fesseau, M. and P. Van de Ven (2014) Measuring inequality in income and consumption in a national accounts framework, Statistics Brief, Oecd, n. 19.

Garofalo G., I. Rocchetti e C. Viviano. 2012. *A revision of the Italian Business Register: a new methodological and conceptual "backbone" for a new informative system on employment*. Washington, D.C. 17–20 September 2012: 23rd Meeting of the Wiesbaden Group on Business Registers.

Garofalo G. 2014 *Il Progetto ARCHIMEDE obiettivi e risultati sperimentali*, Istat working papers n.9/2014

Mc Cully C. P. (2014) *Integration of Micro and Macro Data on Consumer Income and Expenditures* in D.W. Jorgenson, S. Landefeld and P. Schreyer (eds.), Measuring Economic Sustainability and Progress: NBER Book Series Studies in Income and Wealth.

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

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

Ruggles N. D. and R. Ruggles (1974) A Strategy for merging and matching microdata sets, Annals of Economic and Social Measurement, 3 (2)

Ruggles N. D. and R. Ruggles (1975) The Role of Microdata in the National Economic Accounts, Review of Income and Wealth 21, (2).

Ruggles N. D. and R. Ruggles (1987) The Integration of Macro and Microdata for the Household Sector, Review of Income and wealth, 33 (1).

Singh, A.C., Mantel, H., Kinack, M. and Rowe, G. (1993) "Statistical matching: Use of auxiliary information as an alternative to the conditional independence assumption". *Survey Methodology*, **19**, pp. 59–79.

Stiglitz Commission (2009) Report on the measurement of economic performance and social progress, www.stiglitz-sen-fitoussi.fr

Stoker (1993) Empirical approaches to the problem of aggregation over individuals, Journal of Economic Literature, 31(4).

UNECE (2011); *Canberra Group Handbook on Household Income Statistics; Geneva*; http://www.unece.org/fileadmin/DAM/stats/groups/cgh/Canbera_Handbook_2011_WEB.pdf