STATISTICAL MATCHING OF SURVEY DATA FOR THE ANALYSIS OF SPENDING PATTERNS AND POVERTY

Maria Rosaria Ferrante

BRIEF INTRODUCTION
The first aim of this research project is to realize a probabilistic matching among data sets on households’ income, consumption, lifestyle habits and perceived difficulties in their daily life. The Italian National Statistical Institute (ISTAT) releases the following survey data: EU-SILC (Statistics on Income and Living Conditions), HBS (Household Budget Survey) and ADL (Aspects of daily life). Secondly, by taking advantage of the joint availability of households’ micro-data on these dimensions, this research will focus on the evaluation of household well-being, on the analysis of consumer behavior on the low/high income groups, on the evaluation of the impact of the economic crisis on demand for different expenditures categories and income classes, and on the estimation of territorial disparities in well-being.

BACKGROUND AND STATEMENT OF THE PROBLEM
In times of recession it is especially important to understand the multidimensional linkages between income and consumptions, how both determines households’ well-being and how they are distributed across socio-economic groups and territories. In 2008 the French government, in order to identify new tools to measure economic performance and social progress, founded the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz, Sen and Fitoussi, 2010). The Commission has stated that in order to measure well-being of people times are ready to shift the attention from the economic production to the measurement household income, consumption of both goods and time, wealth, and the evaluation of non-market activities. Even if this goal calls the national statistical institutes to produce new household information, the setting up of new surveys is almost unfeasible because of budget constraints and an excessive reporting burden on respondents. This growing demand for data on households’ economic well-being at the micro level is encouraging the production of integrated databases on household income, consumption, habits and the daily difficulties faced by households. Considering this framework, the opportunities offered by the statistical matching of survey data, that is the matching of information coming from different databases, is becoming increasingly popular, and the literature on statistical record linkage is becoming large. Besides, studies on this field have been stimulated by the increasing availability of freely accessible databases, either from administrative sources or from institutional sample surveys. In Italy, information on household income is provided by the EU-SILC survey (Statistics on Income and Living Conditions), information on household consumption is provided by the HBS (Household Budget Survey) and information on household lifestyle habits is provided by the Aspects of daily life (ADL) survey. All these surveys are conducted by ISTAT (Italian National Institute of Statistics), they are freely available to researchers, and they provide relevant information, but separately and for different household samples. The same is true for many other EU countries.

In order to obtain joint observations of all the characters of interest, or, in other words, combine these information for the same household, ISTAT started a project of integration of household income, consumption and wealth (Donatiello et al., 2016), as part of the national modernization program of social statistics. Istat has set up an extensive process of harmonization of national social surveys, and a strong effort has been done in relation to the definitions and classifications of the common variables of the HBS and EU-SILC surveys, in order to achieve those essential pre-conditions for data matching purposes and micro-integration. Istat has investigated the use of the traditional statistical matching techniques as non-parametric imputation (Singh et al., 1993) and methods dealing with complex surveys (Renssen, 1998). Dalla Chiara et al. (2016) also generated an integrated data base on income, consumption, time use (Istat) and data on household conditions (Centro Internazionale di Studi sulla Famiglia) in order to evaluate individual and social welfare and multidimensional poverty and inequality. The author adopt a propensity score matching approach (Rosebaum and Rubin, 1983) and report tests aimed at
evaluating the quality of the procedure at a high level of detail, based on statistical and economic considerations. 

The availability of a data set that contains individual (household) information on consumption, income and habits would allow to deepen the analysis of connections among these phenomena and then to give a multidimensional measure of the household well-being. 

To date, Istat has not yet released a matched household-level data base with such set of information. Thus, the literature on household well-being tends to adopt a macroeconomic point of view and consider, for a variety of aggregated well-being dimension (twelve for Italy), evaluated at the national or subnational level, hence neglecting the microeconomic analysis of this phenomenon (Istat, 2016).

The literature that exploits individual observations on consumption and income, mainly focuses on (a) the link between consumption and income inequalities; (b) the evolution of consumption in relation to trends in income inequality; (c) the effect of income shocks on consumption choices; and (d) the opportunity of using consumption versus income information to detect and measure poverty (Attanasio and Pistaferri, 2014; Attanasio et al., 2015; Blundell et al. 2008; Blundell and Pistaferri, 2003).

Finally, the analysis of territorial disparities in Europe is based on the indicator, “people at risk of poverty or social exclusion”, that is the proportion of the population which is: a) at risk of poverty after social transfers and/or b) severely materially deprived and/or c) living in households with zero or very low work intensity. This indicator, born with the aim of providing a statistical measure of the intangible aspects of poverty and exclusion from the labour market, would be suitably complemented with an indicator of multidimensional poverty calculated at the sub-national level. Small Area Estimation (SAE) methods could be fruitfully be used for this purpose. Currently, SAE models (Rao and Molina, 2015, 2016) mainly focus on unidimensional indicators.

**RESEARCH QUESTION OR HYPOTHESIS, AIM, OBJECTIVES AND DELIVERABLES**

The first aim of this research consists in a literature review on the matching of consumption and income data bases. The literature on matching methods is large, and various attempts at linking income and consumption household data have been taken. This research will draw from these results, and compare them with more recent proposals on statistical matching (Conti et al., 2015, Ahfock et al., 2016; Harron et al., 2016; Tancredi and Liseo, 2011). This first analysis will lead to the choice of the most suitable matching method, hence to the matching of the EU-SILC, HBS and ADL information. The study should benefit from some innovations introduced in social surveys at the European level, which are expected to facilitate the matching process. The Italian HBS has switched to a new expenditure survey in 2014, with a first data release in 2015. The new HBS has been redesigned to harmonize as much as possible the variables in common with the SILC survey. The current revision of the EU-SILC legal basis represents another chance for enhancing the micro integration of social surveys and for broader integration purposes. The future availability of information on consumption and wealth in the 2017 SILC module could markedly improve the quality of the matching results.

The second part of this research will focus on the statistical analysis of the joint distribution of different dimensions of well-being, beyond the separate analysis of income and consumption.

At the moment there is no data source that provides joint information on all the relevant variables for households or individuals. The availability of individual information on income, expenditures, lifestyle habits and perceived difficulties in the household daily life will allow to depict an all-round picture on household well-being, whereas considering these dimension separately does not suffice to provide an appropriate measure of the household standard of living.

In June 2010, the European Council adopted a social inclusion target as part of the Europe 2020 Strategy: to lift at least 20 million people in the EU from the risk of poverty and exclusion by 2020. To monitor progress towards this target, and considering the multidimensional nature of poverty and social exclusion, this indicator consists of three sub-indicators: i) at-risk-of-poverty (i.e. low income); ii) severe material deprivation; and iii) living in very low work intensity households.

With respect to the current situation, the availability of the matched data set would allow to compare people’s exposure to poverty and social exclusion by using different measures: income, expenditure, material deprivation, life difficulties, etc.
Below, some of the analysis made possible by the creation of this new set of information are listed.

**Analysis of the connection among the different information currently used to measure separately aspects of life standard.**

**Measure of well-being in a multivariate dimension.**

For example, this research could answer the following research questions:

- Are those people that are classified as income-poor based on the Eurostat definition of poverty, also expenditure-poor people?
- Are people classified as income-poor also deprived of basic needs?
- Do income-poor people experience higher difficulties or different difficulties in their everyday life relative to non-poor people?

**Does the poverty intensity (income distance from the poverty line) affect the consumption patterns of the poor? And their saving behavior?**

The analysis of the spending patterns of low/high-income households will show how poor/affluent people allocate their income across consumption categories. For each of these groups, this study aims to estimate how an increase in income is allocated across consumption categories. A relevant classification in this context is the one between permanent and transitory poverty.

Finally, we aim to explore the use of savings and credit across poor/affluent household.

**Differential effects of the global recession on consumer behavior of poor/affluent household**

The global recession which followed the 2008 financial crisis has resulted in lower incomes and higher unemployment across Europe, hence affecting consumption behaviors. However, the ultimate effect on consumer choices is yet to be quantified, especially when the focus is on socio-economic disparities, which have been themselves affected by the crisis. The availability of joint information on income, consumption and lifestyle habits allow us to evaluate the impact of the crisis on demand for the various expenditures categories, and whether this effect is different between different income groups.

**Territorial disparities in well-being**

In this framework, the aim is the estimation of a multidimensional poverty rate, cross-classified for sub-national regions and socio-demographic groups. The availability of regional indicators based on dimensions that go beyond income could be of interest to define territorial social policies. The sample size of the social Istat surveys is not sufficient to obtain reliable estimates of regional poverty indicators. The variability of disaggregated direct estimates should be first obtained. Then, small area models, where the multidimensional poverty indicator is the outcome variable, can be proposed.

Obviously, all the above proposed research lines have to take into account the new source of uncertainty implicit in inferential procedures based on probabilistically linked data.

The main **expected results** (leading to the education plan) are the following:

- review of the literature and a critical evaluation of the existing matching solutions and of the methods more recently proposed to link survey data
- realization of the matched data base
- statistical data analysis, the estimation of impact evaluation models and of small area models in order to achieve interpretative results about the economic phenomena along the above directions
- realization of the appropriate statistical codes;
- realization of at least one scientific paper

Since the availability of the matched data set offers the possibility to extend the research in many directions, the first year research activity might lead to the renewal of the research grant for a second year.
**PARTICIPANTS IN THE STUDY AND THE ROLE THEY PLAY**

The successful applicant will be part of the Economic Statistics research group of the University of Bologna, and is expected to foster collaboration within the group, and exploit synergies with other research themes covered by all economic statisticians in the Department, including evaluation methods for quasi experimental data, modelling of consumer behavior, and marketing and consumer research.

On this research theme, existing collaborations at national level include the Department of Economics of the University of Verona (Prof. Federico Perali), with theoretical and applied expertise in matching data-sets from several sources. At the international level, relevant collaboration on the themes of poverty and consumption patterns exist with the Toulouse School of Economics (Prof. Vincent Requillart), with the health economics group of the London School of Economics and Tropical Medicine (Prof. Richard Smith) and – in relation to income inequalities in developing countries – with the School of Oriental and African Studies in London (Prof. Bhavani Shankar).

**REFERENCES**


Harron K., Goldstein H., Dibben C. eds. (2016), Methodological Developments in Data Linkage, John Wiley & Sons, Ltd.


