# EUROPEAN
# POLICYBRIEF

**Small Area Methods for Poverty and Living Condition Estimates**
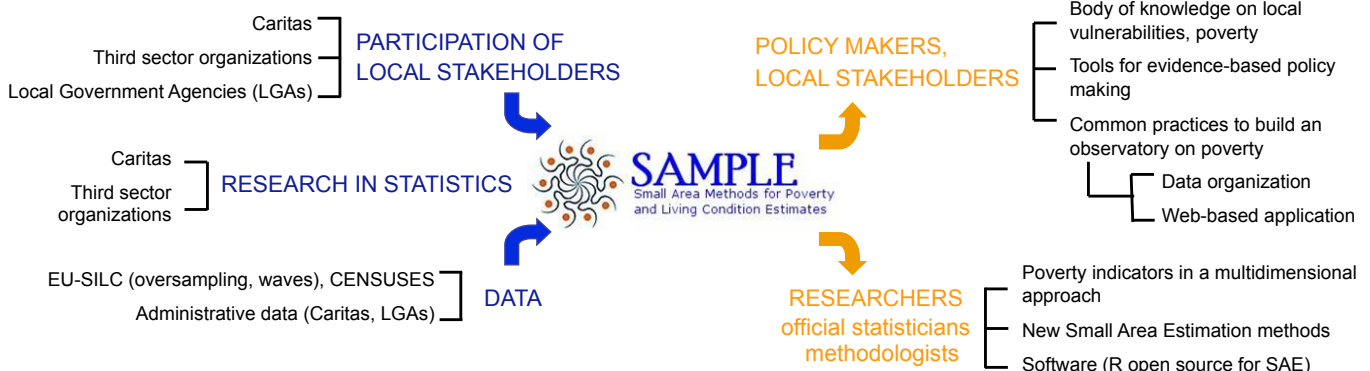
Policy Implications of SAMPLE, an EU-funded research project involving seven public and private research centres and a Public Administration for the research and the estimation of small area poverty indicators

**February 2011**

INPUTS and ACTORS

Caritas
Third sector organizations
Local Government Agencies (LGAs)
PARTICIPATION OF LOCAL STAKEHOLDERS

Caritas
Third sector organizations
RESEARCH IN STATISTICS

EU-SILC (oversampling, waves), CENSUSES
Administrative data (Caritas, LGAs)
DATA

**SAMPLE** Small Area Methods for Poverty and Living Condition Estimates

ACTORS and OUTPUTS

POLICY MAKERS, LOCAL STAKEHOLDERS
Body of knowledge on local vulnerabilities, poverty
Tools for evidence-based policy making
Common practices to build an observatory on poverty
Data organization
Web-based application

RESEARCHERS official statisticians methodologists
Poverty indicators in a multidimensional approach
New Small Area Estimation methods
Software (R open source for SAE)

## INTRODUCTION

**The project**

Given the growing social, demographic and economic problems, the research community, policy makers and practitioners place great emphasis on the development of efficient, effective and reliable indicators and on the collection of high quality data on life conditions, not only at national level but also at regional and at lower geographical levels. In this context, SAMPLE aimed at identifying and developing new indicators and models that will help the understanding of inequality and poverty with special attention to social exclusion and deprivation.

Furthermore, the project developed models and implemented procedures for estimating these indicators and their corresponding accuracy measures at the level of small area (LAU1 and LAU2 level), in order to offer to local, national and European governments accurate data able to: i) ensure monitoring of poverty and inequality; ii) focus their policies on segments of population at higher risk of poverty, some of them specially elusive; iii) appreciate the multidimensional nature of

poverty and inequality with attention to the non-monetary aspects of it, such as social exclusion, vulnerability and deprivation; iv) measure the subjective aspects of poverty as they are perceived by local groups and populations.

## SCIENTIFIC EVIDENCE AND ANALYSIS

**The involvement of third sector organizations: proposal of new indicators**

SAMPLE project previews an important role for local non-profit organizations: we acquired and integrated their databases with administrative databases and we relieved with a special survey (delphi method) their point of view on poverty and social exclusion and on the usefulness of poverty indicators. All data and indicators were considered referring to a multidimensional approach.

The 68,2% of local stakeholders considers that indicators are very useful for the planning of social policies. The 42,1% thinks that they are also very usefulness for the realisation of their activities. They suggest also new indicators to monitor poverty: the debt is the indicator more relevant, the quality of food, the quality of housing. They emphasize also the difficulties in payment of utility bills due to low incomes and the phenomenon of job insecurity. Another indicator suggested is the capability of access to services, which is the knowledge and usability of services to citizens in distress.

Other indicators are those created by the Tuscany Regional Network of Social Observatories to build the Health's Profile at local level: demographic profile, economic profile, health state, elderly persons, families and young, Immigration, disability, mental health, dependences. With SAMPLE project we created a link with this regional activity and we selected all the indicators that have a connection with the phenomena of poverty and social exclusion (most of all).

**Applying Small Area Methodologies to the Italian survey EU-SILC 2007: some results**

DSMAE and SOTON/MANCHESTER have developed small area methodologies based mainly on M-quantile models. In particular, they proposed linear, robust, nonparametric and geographically weighted M-quantile regression models. These methodologies have been applied to obtain estimates at provincial and municipality levels (LAU1 – LAU2) using data from the Italian survey EU-SILC 2007. They focused on three Italian regions: Lombardia, in the North of the Italy, Toscana, in Central Italy, and Campania, in Southern Italy. The choice of these three regions, out of the 20 existing regions in Italy, is motived by the geographical differences characterizing the Italian territory. In particular, DSMAE and SOTON/MANCHESTER have investigated the so-called "north-south" divide characterizing the Italian territory, since each of the three regions can be considered as representative of the corresponding geographical area of Italy (Northern, Central and Southern/Insular). The main result is the higher incidence of poverty in the provinces of Campania. For this region the estimates of the incidence of poverty i.e. the percentage of households below the poverty line (9504 Euros, corresponding to the 60% of the national median income), are between 25% and 44%. Unlike Campania, in Lombardia and in Toscana the ranges of the incidence of poverty are between 10% and 19% and 11% and 26% respectively. Moreover, the estimated mean equivalised income also suggests a gap between these three Italian regions and their provinces.

**Applying Small Area Methodologies to the Spanish surveys EU-SILC 2004-6: some results**

The small area methodologies developed by UC3M and UMH have been applied to data from the Spanish surveys EU-SILC 2004-6, focusing on the whole set of the Spanish provinces. The aim of the applications is to estimate poverty indicators by province and sex. A wide variety of procedures have been applied, including Empirical Best Prediction (EBP) and Empirical Best Linear Unbiased Prediction (EBLUP) method based on time, space and spatio-temporal mixed models. From the socioeconomic point of view, UC3M and UMH have investigated the so-called "North-East to South-West" division characterizing the Spanish territory. From the mapping of poverty indicators at province level, the Spanish regions could be classified in three subsets. First one contains the regions with the lowest poverty incidence, like Cataluña, Aragón, Navarra, País Vasco, Cantabria and Baleares. The second subset contains the regions having an intermediate position, like Galicia, La Rioja, Castilla León, Asturias, Comunidad Valenciana and Madrid. Finally the third subset contains those regions with higher poverty incidence, like Andalucía, Extremadura, Murcia, Castilla La Mancha, Canarias, Canarias, Ceuta and Melilla.

**The 2008 EU-SILC oversampling for the Italian Province of Pisa – Some results**

DSMAE estimated the mean and median equivalised household income, the rate of households declaring to be unable to face unexpected financial expenses and other poverty indicators using data coming form the EU SILC 2008 oversampling for the Province of Pisa. All these estimates are computed at provincial level, that is for the Province of Pisa, and also at a finer geographical level, namely for the five 'Health Societies' of the Province of Pisa.

The mean of the equivalised household income was equal to 18,820 Euros, in the Province of Pisa in 2007. In terms of income quantiles, the 20% of the households in the Province had an equivalised income under 11000 Euros, the 50% under 16707 Euros and the 80% under 23,576 Euros. As far as concerns some of the main household characteristics, in the Province of Pisa an increasing level of education of the head of household corresponds to a higher level of household income, both in terms of mean and percentile estimates. In terms of gender, if the head is a male, the estimated mean household income is significantly higher, around 19,500 Euros, with respect to the nearly 15,500 Euros estimated for the households where the head is a female. The Head Count Ratio (HCR) or *at-risk-of-poverty-rate* of the Province of Pisa is equal to 15.8%. The HCRs estimated for four out of the five Health Societies in the Province are very similar to the provincial estimate, even if characterised by higher estimated standard errors.
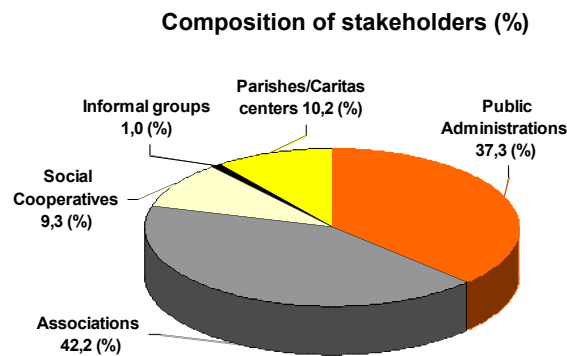
The main result standing from the oversampling direct estimates is that the economic, poverty and social benefits indicators of the five Health Societies of the Province of Pisa are characterized by certain variability. Thus, computing the direct estimates only at provincial level would mask the important differences emerging when repeating the analysis at a more detailed geographical level.

Another interesting result is the variability between the different computed indicators. In particular, areas characterized by low mean and median household income estimates can be characterized by low estimated discomfort indicators. That is, direct income estimates and indicators of perceived economic discomfort can give different indication of the poverty and living conditions in a given area. Thus, it is important to always consider both types of indicators when analysing the areas of interest.

**The local Observatory on poverty**

With SAMPLE project we created a first local Observatory on poverty, vulnerability and social exclusion. The focal points are:
- a formal agreement between the Province of Pisa and three local Caritas Agencies for a permanent monitoring of poverty and social exclusion accessing to the data of Caritas Counselling Centers. Caritas is one of the most important international organisation that realizes actions to contrast poverty
- the involvement of 252 stakeholders (institutions and third sector organisations) to create a local network of qualified "antennas" on the territory. The main instrument for their involvement has been a survey with Delphi Method (http://www.sample-project.eu/it/the-project/deliverables-docs.html see D10 "Final Report task 1.4) and the participation at the development of the software

**Composition of stakeholders (%)**



**A data-centered social application**

Within SAMPLE we developed a web-based software implementing some of the main deliverables of the project. It will feed the local policy makers with robust, disaggregated and up-to-date indicators for a knowledge-based planning. This web application (portal) is meant to be the entry point for social inclusion activities in the province of Pisa. In this perspective, we involved from the beginning all the local stakeholders in the building of the application, asking them for suggestions about the desired software layout and functionalities.

At the end of the project, the software will be used by the Local Observatory on Poverty and Social Exclusion to monitor Laeken indicators and other relevant selected social indicators at province (LAU1) and municipality (LAU2) level. The aim is to improve the knowledge of the phenomena and the local social policies, increasing participation and collaboration between all the local stakeholders involved in fighting and preventing social exclusion. The software allows to store and to update the raw administrative data and the automatic processing and calculation of most indicators. The core engine of the software uses some advanced R functions and SAE algorithms, which improve the quality of the estimations allowing the calculation of poverty indicators. The third module (front-end) disseminates and shares the data and the indicators by creating information-rich, easy to use, easy-to-understand graphics (unveiling the meaning of data without hiding their complexity). The front-end is empowered by social tools (blog, feed, personal profiles, etc.) aiming to boost the public awareness and the participation between local stakeholders. → http://lab.sample-project.eu[1]

---

[1] Temporary beta address. The final address will be decided after the conclusion of a last stakeholders poll (off-project)

## POLICY IMPLICATIONS AND RECOMMENDATIONS

**Local policymaking: Bridging the gap between research and policies**

SAMPLE project developed a direct channel of communication between a research organisation (University of Pisa) and a Public Administration (Province of Pisa). The result is a direct connection between the results of the research and the need of policy planning. The SAMPLE software is the instrument that realizes this connection and also the effort to achieve a common language on indicators and methods and the growth of common practices of research and data acquisition.

**Indicators for social policies planning**

With SAMPLE project we developed – in collaboration with Tuscany Region – a set of indicators that will be used by the Health's Societies for their planning of social policies. This set included some indicators from Revenue Agency data, from Labour Market data and from many other relevant administrative social and health databases. SAMPLE software collects this data and returns a social profile of the territory, which is very useful for policymakers. The Province of Pisa stipulated a protocol with one of the five Health's Society that previews the collection of indicators every year.

**Bring statistics to the people**

An effective policymaking needs data. At local level, there's a great lack of data pinpointing the differences between areas and social groups. It's often difficult to get access to accurate data and reliable statistics. Also, often for the local policy makers is not so easy to understand complex indicators and "silent" raw data. SAMPLE tries to fill these gaps. The web application has been planned following the most recent trends and best practices in the field of statistical data visualisation and data sharing. New technologies, born outside of the traditional statistical agencies, are giving people plenty of new opportunities for uploading and manipulating statistical data. That's what we will try to do with the web application: serving the larger community raw indicators and statistics, and give them the means to reuse them with creativity, share them and make them "living". In a statement, "bring statistics to the people". *"If people put data on the Web - government data, scientific data, community data - whatever it is, it will be used by other people to do wonderful things in ways they never could have imagined."* (Tim Berners-Lee)

**Recommendations:
The issue of privacy policies**

The important issue of privacy policies applied on available data in Italy and also in EU state members affected the project implementation and forced us to overcome many obstacles, to acquire the dataset previewed by the project. So, referring to our experience, these are some suggestion for this kind of projects:

- to involve, like official partners, the more important Public Administrations that have relevant data for the project. This could avoid many obstacles of the law.
- to promote formal agreement with other Public Administration for having access to data that are useful for the project and for policy planning.

**Recommendations: the different souls of SAMPLE project**

We needed to harmonise different languages, different approach and point of view of social phenomena that characterise the University and Public Institutions. This difficulty has required a major effort in terms of meeting, discussion and continuous exchanges between the partners.

Policymakers have been involved in the process of building new indicators since the extent to which new indicators and research results, even crucial, influence policies and decisions is conditioned by the day-by-day current working practices of local agents

## RESEARCH PARAMETERS

**New multidimensional poverty indicators**

The contribution by CRIDIRE, WSE and GUS aimed at identifying and developing new indicators and models that will help the understanding of inequality and poverty. They appreciate the multidimensional nature of poverty and inequality with attention to the non-monetary aspects of it, such as social exclusion, vulnerability and deprivation (see Report Sarkozy, Sen 1985); poverty is measured along its subjective aspects as they are perceived by local groups and populations. These indicators refer to the definition of fuzzy measures of the most important indicators of objective and subjective well-being, i.e. measures of satisfaction and happiness that are good proxy of individual welfare. At this end, the official European data from the Survey on Income and Living Conditions (EU-SILC) has been the main source of data. The indicators have been constructed using data from several waves (called pooled estimates) and the results have been compared between different EU-SILC waves with focus on the local longitudinal changes

**Small Area Estimation**

The most popular approach to small area estimation employs random effects models for estimating domain specific parameters. Empirical Best Linear Unbiased Predictors (EBLUP) are widely used when population parameters are linear functions of the target variable. If those parameters are nonlinear then more elaborated techniques are required. When modelling at the unit-level UC3M has developed the so called Empirical Best Unbiased Predictor approach and UMH has proposed to use the nested structure of the sample and population to apply unit-level models at a low aggregated level.

An alternative solution to the estimation of non linear parameters (like the poverty indicators) is to use area-level models. In this framework UC3M and UMH have introduced EBLUPs based on time, space and spatio-temporal models. They also have given statistical techniques to estimate the mean squared errors and they have implemented the corresponding software.

An alternative approach to small area estimation that relaxes the parametric assumptions of random effects models employs the M-quantile models.

Having developed the methodology for estimating small area averages and quantiles by M-quantile models, DSMAE and SOTON/CCSR focus on the estimation of poverty indicators, which present a special case of estimating small area quantiles. they consider estimation for two popular poverty indicators namely, the Head Count Ratio (HCR) and the Poverty Gap. In addition, they also consider estimation for fuzzy set indicators that have more recently attracted interest in poverty studies.

DSMAE has extended the p-spline regression model to the M-quantile method for the estimation of the small area parameters using a nonparametric specification of the conditional M-quantile of the response variable given the covariates. This is useful when the functional form of the relationship between the response variable and

the covariates is unknown or has a complicated functional form.

Finally when the observations that are spatially close may be more related than observations that are further apart DSMAE and SOTON/CCSR have proposed an M-quantile Geographically Weighted Regression models that take into account the spatial correlation by assuming that the regression coefficients vary spatially across the geography of interest

## Integration of EU-SILC data with administrative and third sector database

Local government needs reliable statistics in order to address policies and actions against poverty and social exclusion. Generally, local government agencies have a huge amount of administrative data, which can help calculating statistics. However such administrative data are seldom integrated in order to convey a synthetic view on poverty, deprivation and social exclusion. For this purpose data integration procedures have been set up in order to match sample and administrative data at local level. Among the main advantages, there is the possibility of grossing up administrative indicators using EU-SILC sample weights. Additionally, the matched data sets allow to give interpretation of administrative indicators taking into account the characteristics of the households whom the individuals belong to. On the contrary, relying on administrative indicators only, individuals may be classified as poor or emarginated, in spite of actually living in a wealthy family.

The application benefited from the EU-SILC survey oversampling for the Italian province of Pisa (PI-SILC in the following) and from the access to administrative databases not frequently used for statistical analysis. Unfortunately, due to confidentiality issues, we could actually access less than the expected amount of administrative data sets. Moreover some demographic variables were removed form such data sets making our task even more complicated. The available administrative databases were: i) The Caritas database ii) The Revenue Agency database (RA); iii) the provincial Job Centre database (JC).

Caritas database collects data on people hardly detectable in the PI-SILC sample, being more often homeless people. For this reason the attempt to match PI-SILC and Caritas observations resulted in very few matched units. Caritas data were thus discarded and only the RA and JC databases were used in the integration procedures.

On the contrary, the RA and JC data sources seem to cover populations similar to the SILC population. As a consequence we expect to find some of the PI-SILC sampled individuals in the administrative archives. The exact matching (or records linkage) is the technique used to identify and pick up such units.

Records 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 A and B files are supposed to contain identical units that have to be found according to an identifier (like the social security number) or a set of identifying variables (k variables) present in both files. The procedure is very simple provided that each record in both files contains the same identifier and this identifier is recorded without errors. In this case the problem is solved by simply picking out the records (if any) with the same identifier value. This procedure is known as exact matching.

Unfortunately, some complications may occur: (i) 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. (ii) Some values of the *k* variables may be missing so that the *k*-variable may not be known exactly for some of the records in *A* or *B*.

Formalizing the linking procedure into a statistical model, it is possible to evaluate the quality of the matching by measuring the probability of generating false-matched-pairs and false-unmatched pairs.

The output of the integration procedure is a dataset where the PI-SILC records are extended with variables taken from the RA and JC databases. Furthermore a linking probability is provided which helps evaluating the quality of the matching. This is the probability that the coupled units represent a true link (the two records refer to the same unit) giving the observed values for the matching variables.

The results of integration are not entirely satisfying in that only a limited number of PI-SILC units finds at least a link in the administrative databases. This is mainly due to confidential issues which prevented us from gaining access to personal identifiers in the administrative databases. One of the finding of the project is that it is extremely difficult to access administrative databases for research purposes even when these databases are accessible to the project's partners.

## PROJECT IDENTITY

| | |
|---|---|
| **Coordinator** | Prof. Monica Pratesi, Università di Pisa - Dipartimento di Statistica e Matematica Applicata all'Economia (UNIPI-DSMAE) |

| | |
|---|---|
| **Consortium** | • Università di Siena - Centro Interdipartimentale di Ricerca sulla Distribuzione del Reddito (CRIDIRE), ITALY<br>*Prof. Achille Lemmi* lemmi@unisi.it<br>• Cathie Marsh Centre for Census and Survey Research, University of Manchester (CCSR), UK / School of Social Sciences University of Southampton (SOTON), UK<br>*Dr. Nikos Tzavidis* N.TZAVIDIS@soton.ac.uk<br>• Departamento de Estadística, Universidad Carlos III de Madrid (UCM3), SPAIN<br>*Dr. Isabel Molina* isabel.molina@uc3m.es<br>• Centro de Investigación Operativa, Universidad Miguel Hernandez de Elche (UMH), SPAIN<br>*Prof. Domingo Morales* d.morales@umh.es<br>• Warsaw School of Economics (WSE), POLAND<br>*Prof. Tomasz Panek* tompa10@interia.pl<br>• Provincia di Pisa - U.O. Studi e Ricerche - Osservatorio per le Politiche Sociali - Ufficio Politiche Comunitarie (PP), ITALY<br>*Dr. Paolo Prosperini* p.prosperini@provincia.pisa.it<br>*Dr. Claudio Rognini* c.rognini@provincia.pisa.it<br>• Simurg Ricerche (SR), ITALY<br>*Dr. Moreno Toigo* simurg@simurgricerche.it<br>• Central Statistical Office (GUS, Główny Urząd Statystyczny), POLAND<br>*Dr. Mariusz Kraj* m.kraj@stat.gov.pl |

| | |
|---|---|
| **EC contact** | Dr. Ian Perry, Ian.Perry@ec.europa.eu |

| | |
|---|---|
| **Funding scheme** | **SEVENTH FRAMEWORK PROGRAMME**<br>**THEME FP7-SSH-2007-1**<br>*Socio-economic sciences and the Humanities Part 8*<br>Collaborative Project **-** Small or medium-scale focused research project<br>Grant Agreement n.: 217565 |

| | |
|---|---|
| **Duration** | March 2008 – February 2011 (36 months) |

| | |
|---|---|
| **Budget** | EU contribution: € 774.972,00 |

| | |
|---|---|
| **Website** | www.sample-project.eu |

| | |
|---|---|
| **For more information** | coordinator@sample-project.eu |