



SAMPLE DELIVERABLE 26

Third Mini-Project progress report

Grant agreement No:	SSH - CT - 2008 – 217565
Project Acronym:	SAMPLE
Project Full title:	Small Area Methods for Poverty and Living Conditions Estimates
Funding Scheme:	Collaborative Project - Small or medium scale focused research project
Deliverable n.	26
Deliverable name:	Third Mini-Project progress report
WP no.:	5
Lead beneficiary:	7
Nature:	Report
Dissemination level:	PU
Due delivery date from Annex I:	31/08/2009
Actual delivery date:	23/12/2009
Project co-ordinator name:	Mrs. Monica Pratesi
Title:	Associate Professor of Statistics - University of Pisa
Organization:	Department of Statistics and Mathematics Applied to Economics of the University of Pisa (UNIFI-DISMAE)
Tel:	+39-050-2216252, +39-050-2216492
Fax:	+39-050-2216375
E-mail:	coordinator@sample-project.eu
Project website address:	www.sample-project.eu



Third Mini-Project progress report

Project Acronym: SAMPLE
Project Full title: Small Area Methods for Poverty and Living Conditions Estimates
Grant agreement No: SSH - CT - 2008 – 217565
WP no: 5
Deliverable n. 26
Document title: Third Mini-Project progress report
Status: Final
Editors: Marta Garro, Angela Colucci, UNIPI DSMAE
Authors: Achille Lemmi (lemmi@unisi.it), WP1 leader, CRIDIRE
Domingo Morales (a.coli@ec.unipi.it), WP2 leader, UHM
Claudio Rognini (c.rognini@provincia.pisa.it), WP3 leader, PP-UROPS
Moreno Toigo (simurg@simurgricerche.it), WP4 leader, SR
Paolo Prosperini (p.prosperini@provincia.pisa.it), WP5 and WP6 leader, PP-PC
Document Editing: Marta Garro, Angela Colucci
Language revision: Marta Garro

SAMPLE WP1 activity (1 March – 31 August 2009)

The main targets of WP1 are to analyze the mechanisms and the determinants of poverty and inequality and to translate them into effective indicators. CRIDIRE-Siena University have dedicated the third semester of the project mainly to develop new multidimensional and fuzzy measures of poverty and re-sampling methods for variance estimation for these measures. In particular, a new approach (Integrated Fuzzy and Relative Approach) has been introduced: it combines the Totally Fuzzy and Relative Approach of Cheli and Lemmi (1995) and the approach of Betti and Verma (1999). A factor analysis has been conducted to construct certain dimensions of non-monetary indicators of living conditions. Concerning the re-sampling methods for variance estimation, Jack-knife Repeated Replication (JRR) method has been adopted. The developed methodology has been tested with the first available waves of EU-SILC data. SAS codes have been developed for estimating the new fuzzy and non-fuzzy approach poverty measures and for the corresponding standard errors. Programs from SAS to R for poverty measures and from SAS to R for small area estimates have been partially converted. In this semester, Siena University has started also to produce the first preliminary methodological developments in Pooled estimates of indicators.

The partners of WP1 – WSE and CES-GUS - introduced proposition of including in the final set of indicators two additional fuzzy measures of the depth of relative poverty and deprivation (FMD and FSD). The Integrated Fuzzy and Relative Approach (CRIDIRE) contains indicators which are counterparts of headcount index in traditional approach whereas the Fuzzy Monetary Depth and the Fuzzy Supplementary Depth indicators (WSE) are counterparts of poverty gap index in traditional approach. The methodology of computing two additional indicators and variance estimation using bootstrap techniques has been tested with the last available data from Polish EU-SILC (2007). SAS codes have been developed for estimating Fuzzy Monetary Depth and the Fuzzy Supplementary and for the standard errors for all fuzzy measures (introduced by CRIDIRE and WSE) using bootstrap techniques.

The research group of CRIDIRE, GUS and WSE has met in Siena (3-4 June 2009) and in Warsaw (1 July 2009).

In the same period, UNIPI-DSMAE has been involved in the following activities: i) preparing slides on the over-sampling state of the art for the Elche meeting (which are available on SAMPLE web site); ii) writing Deliverable 7 (the work is still in progress). EU-SILC microdata of over-sampling will be released by the end of December 2009. At present, Istat is completing the data check for missing values.

Simurg has dedicated the third semester of the project for WP1 to carry on the activities started in the first year about task 1.4, jointly with PP-UROPS. The list of indicators to monitor social exclusion developed at the end of 2008, has been widely discussed within the Social Observatories network in several meetings at regional level. Simurg has attended at these meeting and has actively participated at the methodological definition of the regional guidelines in the area of poverty and social exclusion indicators. Simurg and PP-UROPS have also disseminated the SAMPLE experience during these meetings. The Tuscany Region in November will approve specific guidelines for social policies planning that will be addressed to the Health Societies, the public organisation that in Tuscany Region has the competence in social policies. Moreover, Simurg is planning with PP-UROPS the sharing of the indicators with the local public and private stakeholders by the way of a survey conducted on local stakeholders. It has supported PP-UROPS in the development of the methodology and of the tools to be used for the survey.

SAMPLE WP2 activity (March 1 to August 31, 2009)

During the third project semester the partners of WP2 have been mainly involved in the first small area estimation developments. For this sake a lot of research has been done in the main topics that are being developed. In what follows we describe the achieved developments of the project objectives and the work performed during the period March 2009 to 31 August 2009. As poverty indicators are nonlinear on unit-level welfare variables, within tasks 2.2 and 2.3 area-level time and space models have been proposed for constructing Empirical Best Unbiased Predictors (EBLUP). Algorithms to fit the models, to obtain EBLUPs and to estimate their mean squared errors have been developed. Regarding tasks 2.2 and 2.3, an interesting new approach to estimate poverty indicators that uses Empirical Best Predictors has been also developed. For this sake a model is fitted to a conveniently transformed unit-level sample data. Then the distribution of the non observed data, conditioned on the observed data is used to simulate the out-of-sample data. Finally the Monte Carlo method is used to compute the EBP. Within task 2.1, robust methods have been investigated and developed. Special emphasis has been placed on Quantile / M-quantile Models for Small Area Estimation, and on estimation of cumulative distribution functions. This last issue allows estimating quantiles of welfare variables. In addition, a robust version of the EBLUP (Sinha and Rao 2009) has been studied and contrasted against M-quantile-based predictors. For tasks 2.1-2.3, simulation experiments to study the behavior of the new methodologies under different scenarios have been carried out and first applications with real data have been done. All these developments will be largely included in the forthcoming deliverable 8 on “first small area developments”.

Main results derived from the research activities of tasks 2.1-2.3 are described in papers. In what follows these papers are listed in three categories: (1) published papers, (2) submitted papers, and (3) other submitted papers partially funded by SAMPLE.

Published papers

Molina, I., Salvati, N., Pratesi, M. (2009) Bootstrap for estimating the mean squared error of the Spatial EBLUP, *Computational Statistics*, 24,3, 441-459.

Pratesi, M., Salvati, N. (2009) Small area estimation in the presence of correlated random area effects, *Journal of Official Statistics*, vol. 25, 37-52.

Tzavidis, N., Marchetti, S., and Chambers, R. (2009). Robust estimation of small area means and quantiles. To appear in the *Australian and New Zealand Journal of Statistics*.

Molina I. and Morales D: (2009). Small area estimation of poverty indicators. To appear in *BEIO* (Bulletin of the Spanish Statistical Society).

Submitted papers

Chambers, R., Chandra, H., and Tzavidis, N.: Mean Squared Error estimation for linear predictors for domains. Submitted to *Survey Methodology*.

Esteban M.D., Morales D., Pérez A. and Santamaría L. (2009). Small area estimation of poverty indicators under area-level time models. Submitted to *CSDA*.

Giusti, C., Pratesi, M., Salvati, N. (2009) Small area methods in the estimation of poverty indicators: the case of Tuscany. Submitted to *Politica Economica*.

Molina I. and Rao J.N.K (2009). Small Area Estimation of Poverty indicators. Submitted to the Canadian Journal of Statistics.

Other submitted papers partially funded by SAMPLE

Datta G.S., Kubokawa T., Rao J.N.K. and Molina I. Estimation of mean squared error of model-based small area estimators (2009). Submitted to Test.

Salvati, N., Tzavidis, N., Chambers, R. and Pratesi, M.: M-quantile geographically weighted regression for small area estimation. Submitted to Test.

Salvati, N., Chandra, H. , Ranalli, M.G. and Chambers R. (2008). Small Area Estimation Using a Nonparametric Model Based Direct Estimator. Submitted to Journal of Computational Statistics and Data Analysis.

Salvati, N., Ranalli, M.G. and Pratesi, M. (2009) Small area estimation of the mean using nonparametric M-quantile regression: a comparison when a linear mixed model does not hold. Submitted to Journal of Statistical Computation & Simulation.

It is also important to remark that the UMH partner has organized the SAE2009 Conference on Small Area Estimation (June 29 to July 1, 2009), where a lot of collaborative and dissemination activities were carried out. The developments of the SAMPLE project were presented to keynote researches and public statisticians attending to SAE2009. A joint cooperative meeting with the colleagues of project AMELI was organized and a press conference was given. Members of the research teams of the SAMPLE WP2 partners presented their advances in the project in the following conferences: ITACOSM 2009 (June 10-12, 2009), Joint Statistical Meeting 2009 (August 1-7, 2009), International Statistical Institute 2009 (August 16-22, 2009).

Concerning dissemination, several seminars, talks and courses have been given. They are listed below:

Initiative: A talk on M-quantile linear mixed models by partner University of Manchester.
Place and date: University of Pisa, January 2009.

Initiative: A course on Small Area Estimation, by partner UMH
Place and date: Instituto Galego de Estadística (IGE), Santiago de Compostela, Spain. March 26-27, 2009.

Initiative: A talk describing the methodological developments and applications of the SAMPLE project, by partner UMH.
Place and date: University Milano-Bicocca, Milano, Italy. April 16, 2009.

Initiative: A talk describing the methodological developments and applications of the SAMPLE project, by partner UMH.
Place and date: University of Sevilla, Sevilla, Spain. May 15, 2009.

Initiative: A talk on recent methodological developments for the estimation of poverty measures by partner DSMAE.
Place and date: University of Florence, Firenze, Italy. May 30, 2009.

Initiative: A talk describing the methodological developments and applications of the SAMPLE project, by partner UMH.

Place and date: Spring School was organized by the European Centre for Soft Computing in Mieres, Spain. May 19, 2009.

Initiative: A talk describing the methodological developments and applications of the SAMPLE project, by partner UMH.

Place and date: Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic. July 8, 2009.

Initiative: A talk on M-quantile linear mixed models by partner DSMAE.

Place and date: University of Milano-Bicocca, Milano, Italy. July 8, 2009.

Initiative: A course on “Small Area Estimation Methods, Applications and Practical Demonstration” by the project leader of partner UC3M together with J.N.K. Rao.

Place and date: Durban, South Africa. August 14-15, 2009.

SAMPLE WP3 activity (March 1 to August 31, 2009)

The main objective of WP3 is to exploit administrative and third sectors' locally available data to calculate indicators to monitor social exclusion and poverty and, moreover, useful to define effective local social policies.

In the last semester, (M 12-18) partners of WP3 have realised these main activities:

a) Access to INPS database (Task 3.1)

After informal contacts during the first 12 months we made a request to INPS for accessing and acquiring on electronic format data which are contained in record related to the following three internal databases: Database of Positions Active contains workers' data; Database of Pensions contains data of pensions according to amount, pensions supplemented by guaranteed minimum income, etc.; Database of ISEE (Indicator of the Equivalized Household Economic Position) contains ISEE declaration's data. ISEE declaration must be submitted by people who want to obtain reduced tariffs in order to gain access to services provided by Public Administrations (school, social housing, etc...). We are now waiting for the official answer from INPS.

b) Access to Revenue Agency database (Task 3.1)

UROPS gained access to this database. The Revenue Agency uses the SIATEL System which contains data on tax returns from 2004 to 2007. For the Sample Project, UROPS obtained from the Department of Finance the permission to gain access to SIATEL database using a password. After verbal agreements, we requested them the authorisation to export the whole database. At the moment we only have access to individual data.

c) Access to Jobcentre –database

We have organised specific meetings with jobcentres' representatives in order to illustrate the Sample project and request collaboration. The Provincial Jobcentre uses the IDOL System to store data; the IDOL database contains data related to people registered as unemployed and to the start and cessation of jobs provided by companies. The IDOL database also contains information upon income and familiar conditions of unemployed people and upon workers' professional status. People who are stored in the Provincial Jobcentre's database result to be 30.574 at 31/12/2008, a number to which must be added the quota of workers communicated by companies. After these meetings, UROPS has gained access to IDOL database with individual code. The archive will be set in a specific server created by Pisa Province for Sample project.

d) Exploration and analysis of local third sector information systems (Task 3.2)

In these months we have organised specific meetings with the local responsible of the Caritas MIROD network. Afterwards, UROPS has gained the access to the MIROD database and has started a collaboration with the three territorial Caritas in order to formalize the collaboration in Sample project with a formal agreement.

e) Local stakeholders selection and involvement (task 3.4)

In these months partners of WP3 have realized two important activities: we have created an early list of local stakeholders and we have created the questionnaire that will be sent to stakeholders. In order to select stakeholders, we have organised few meetings with the third sector provincial responsible – as a key person - and we have identified more than 200 public and private organisations.

SAMPLE WP4 activity (1 March – 31 August 2009)

In the third semester we carried out some investigations aiming to build a graphical user interface (GUI) for R language. We have finally chosen to build our solution upon the opensource duo PHP language/MySQL database: we will start from an existing project (R-php) developed by the University of Palermo and will extend it with more functionality in the back-end. We have also faced with privacy concern implied in the project, studying the state of the art (i.e. CASC Project, 5th FP, CENEX, 6th FP, ESSNet that have produced tools and handbooks for Statistical Disclosure Control) and getting connected with similar projects (i.e. Amerika project by Milano Municipality, and a project by Livorno Municipality). In this perspectives, we have studied the implications of the “Open Governement Initiative” stated by President Obama’s administration in January 2009 and the debate about the disclosure of administrative and government data. During the Elche meeting we have discussed with the other partners about how we could share the R routines developed in WP1 and WP2. UMH has proposed to the other partners a common framework to document the R routines in a way that they could be correctly understood by the WP4 development team.

Concerning the third module (reporting), we have studied some best practices in the field of statistical data representation (gapminder.org, www.visualizingeconomics.com, manyeyes.alphaworks.ibm.com/manyeyes/, www.city-data.com, www.perceptualedge.com, www.livingwage.geog.psu.edu, and many others). We have also reviewed the main contributions on these topics presented at the NTTS Conference in February 2009 in Brussels.

On the front-end, our application will be a web 2.0 oriented PHP/MySQL website with advanced graphic visualizations made with Adobe Flex technology.

The potential users have been identified during the project development, mainly in WP1.4 and WP3.2. For more details see WP1, task 1.4 and WP3 reports.

Finally, in the third semester the development team of WP4 has been enlarged and established. It is composed by: Moreno Toigo, Simurg scientific responsible, who will coordinate the development, the global architecture and functions of the application; Massimiliano Alù, who will develop the web site structure, and the communication styles of the application; Massimiliano Passaglia, who will develop the DBMS and the core code.

SAMPLE WP5 activity (1 March – 31 August 2009)

For a complete description of management activities, please refer to the SAMPLE Periodic Report.

SAMPLE WP6 activity (1 March – 31 August 2009)

During the last semester the main activities on dissemination were related to implement the web site and try to identify the interconnection with other similar research projects. In particular, the focus on the web site concerned:

1. describing in detail the tasks undertaken by each project partner in an apposite technical section on the web site;
2. enabling the upload of project related reports, dossiers and briefs that describe the results and the relevant methodology;
3. creating an online forum for topic discussion, more institutional like eworsae, esof 2010, CER; and more user friendly like facebook, putting in connection Sample with other similar projects;
4. Planning the editor of newsletters about the implementation of the project, the slant of this tool of communication;
5. providing information about future and past events related to the project;
6. implementing the restricted area.

The work carried out was:

1. redacting and publishing of technical fiche about dissemination and management in the apposite technical section of the web area;
2. creating a specific section on the web to put in evidence this report activities;
3. creating new account and internet page of social and research network;
4. studying a draft and a model of newsletter;
5. implementing the web sections, defining the event agenda, organizing the Elche meeting and planning the future event in Varsavia.

The main results achieved during the period were:

- To complete the public area of the web site;
- To give more visibility on web site of sample contents about methodology and results achieved;
- To give more evidence to the Sample project in social and scientific networks;
- To organize the next project meetings.