Enhancing Small Area Estimation Methods - Applications to Istat’s Survey Data

Michele D’Alò, Stefano Falorsi, Loredana Di Consiglio, Fabrizio Solari - Istat, Italy
Monica Pratesi, Nicola Salvati - University of Pisa, Italy
Maria Giovanna Ranalli - University of Perugia, Italy

The goal of this paper is to analyse the possibility to improve the performances of the estimation at sub-regional levels for Istat surveys. In particular, we refer to small domains cutting across survey strata; e.g. estimation of unemployment rate for Local Labour Market Areas - defined as aggregation of municipalities - currently estimated by means of EBLUP based on linear mixed model with spatially correlated area effects.

In this work we exploit the use of semi-parametric models based on penalized splines to prevent misspecifications of the functional relationship between the target variable and some covariates. Moreover, in the same way, semi-parametric modelling can be used to include more accurate spatial information using low-rank thin plate splines (Opsomer et al, 2008). Finally, we analyse the possibility of reducing bias of EBLUP through the introduction of model based direct estimators (Chambers e Chandra, 2006).

The performances of the suggested methods are reported on simulated experiments on 2001 Census data.

References

A Balanced Sampling Approach for Multi-way Stratification Designs for Small Area Estimation

Piero Demetrio Falorsi, Paolo Righi - Istat, Italy

The present work illustrates a sampling strategy useful for obtaining planned sample size for domains belonging to different partitions of the population and in order to guarantee that the sampling errors of domain estimates are lower than given thresholds. The sampling strategy that covers the multivariate multidomain case is useful when the overall sample size is bounded and consequently the standard
solution of using a stratified sample with the strata given by cross-classification of variables defining the different partitions is not feasible since the number of strata is larger than the overall sample size. The proposed sampling strategy is based on the use of balanced sampling selection technique and on a greg-type estimation. The main advantages of the solution is the computational feasibility which allows one to easily implement an overall small area strategy considering jointly the design and estimation phase and improving the efficiency of the direct domain estimators. An empirical simulation on real population data and different domain estimators shows the empirical properties of the examined sample strategy.

Assessing Uncertainty of the Temporal EBLUP: a Resampling-Based Approach

Luis N. Pereira - University of the Algarve – ESGHT, Portugal
Pedro S. Coelho - New University of Lisbon – ISEGI, Portugal

Large scale sample surveys are usually designed to produce reliable estimates of various characteristics of interest for large geographical regions or subgroups of a population. However, for effective planning in a wide variety of fields, there is a growing demand to produce similar estimates for smaller geographical areas and subpopulations for which adequate samples are not available. In fact, sample sizes are very small or even zero in many small areas of interest, which results in unreliable direct design-based small area estimates. This makes it necessary to employ indirect estimators that “borrow information” from related small areas through linking models, using recent census and current administrative data, in order to increase the effective sample size and thus precision. Such indirect estimators are often based on explicit linear mixed models that provide a link to a related small area through the use of supplementary data. The empirical best linear unbiased prediction (EBLUP) approach is the most popular method for the estimation of small area parameters of interest. But the quality measure of the EBLUP estimators is a challenging problem due to difficulties on estimating the mean squared prediction error (MSPE) of such estimators.

This work assumes that the small area parameters of interest follow a Rao-Yu longitudinal model (Rao and Yu, 1994). The Rao-Yu model is a special case of the general linear mixed model involving autocorrelated random effects with a homogeneous covariance structure [first-order autoregressive plus common covariance: AR(1)+J].

This paper reviews the asymptotic analytical approximation of the MSPE of the temporal EBLUP estimator proposed by Rao and Yu (1994), which is corrected to a second order for a small number of time points and a large number of small areas. Under the model due to Rao and Yu (1994), parametric bootstrap and jackknife procedures are proposed for estimating the MSPE of the temporal EBLUP, using similar ideas as in Butar and Lahiri (2003) and Chen and Lahiri (2005). A simulation study was carried out in order to compare the resampling methods with the asymptotic analytical approximation estimates. Our results show that the bootstrap and jackknife estimators perform very well compared to the Rao-Yu analytical approximation on measuring the quality of EBLUP small area estimates, even for a small number of time points and a large number of small areas. Finally, it is presented an application with real data from the Prices of the Habitation Transaction Survey conducted by the Portuguese Statistical Office.

References
Different Approaches for Evaluation Precision Small Area Model-based Estimators

Francesca Inglese, Monica Russo - Istat, Italy
Aldo Russo - Roma Tre University, Italy

This paper shows an overview of all possible evaluation approaches present in statistical literature, that are useful to select a suitable small area estimator among a number of competing small area models, with reference to different experimental contexts of study. In particular, there are situation in which we have only one small area and we have to choose the best estimator, in this case we use methods that provides with area-specific measures of precision that can be associated with each small area estimate; another context of study is when researcher has a set of small areas and he has to study the performances of a number of estimators, so it is useful to adopt a global measure that averaged over small areas (Rao, 2003). Relatively to the last case, new methods? introduced by G. Brown, R. Chambers, P. Heady and D. Heasman (2001)? are useful to evaluate level of bias of different model-based estimators.

Relatively to the first context of study above mentioned, we suggest to calculate an estimate of Mean Squared Error for each model-based estimator examined, using data derived from the only sample available. For what concern the estimate of the above Mean Squared Error we adopt the methodology suggested by Prasad and Rao (1990).

As regard to global measures, it is necessary, for evaluating empirical properties of the considered estimators, to use Monte Carlo simulation method in which we select R samples, each one with n prefixed unit. Then for each sample we estimate the parameter of interest for each small area, using from time to time all the estimators examined.

Global indicators useful to choose the best estimator are commonly known as values of average absolute relative bias, average relative efficiency and average absolute relative error.

The last methods presented help researcher in the choice of small area model that is not necessarily nested and there is some doubt about the assumptions underpinning the model. In particular, we discuss four methods, named Bias, Goodness of fit, Coverage and Calibration, that are based on the crucial assumption that the direct estimates of the small area values of interest are unbiased and the confidence intervals associated with these estimates achieve their nominal coverage levels. In the first method we plot the model-based estimates as X of a graph and direct estimates as Y for providing a visual illustration of bias and, by comparing the OLS (ordinary least squares) regression line with Y=X, a parametric significance test for the bias; the aim of Goodness of fit method is to check for unconditional bias in the model-based estimates, therefore we use a Wald goodness of fit statistic W to test whether there is a significant difference between the expected values of the direct estimates and the model-based estimates; the third...
method evaluates the validity of the confidence intervals generated by the model-based small area estimation procedure; the last method measures the amount of scaling required to calibrate the small area model-based estimates to the higher level direct estimates.

References

fringles@istat.it