General Linear Models in Small Area Estimation: An Assessment in Agricultural Surveys

Carlo Russo, Massimo Sabbatini and Renato Salvatore

Department Economia e Territorio
University of Cassino, Italy
russocar@unicas.it, m.sabbatini@eco.unicas.it, rsalvatore@unicas.it

In recent years one of the most difficult problem to solve by the official statistical institutes is the information delivery about small geographical areas. In the context of agriculture, there is a growing demand for reliable small area statistics in order to assess or to put into effect agricultural policies and programs. Data collected form some well-organized sample surveys, like the farm structure surveys, can be used only to derive direct estimates for large areas, such as administrative districts or regions of national territories. Sample sizes in small domains of interest are typically reduced by the survey design itself, if the objective is to provide accurate estimates at higher level of aggregation than that of small areas. In any case, direct estimates in specific areas are not reliable, because the smallness of sample sizes in the areas can drive to unacceptably large standard errors.

One of the most recurrent method of small area estimation is based on models that consider composite estimates (Ghosh and Rao, 1994). This models combine synthetic and direct estimates, balancing the potential bias of the synthetic estimates with the instability of the direct estimates by a weighted average of the two functions. Some models (Ghosh and Rao, 1994; Rao, 2002) such as EBLUP (empirical best linear unbiased prediction), involve design-based and model-based random variables, and analyze random area effects through the use of area level and unit level mixed linear statistical models (Robinson, 1991). The linear regression models link functions of continuous response variables of interest with some area level covariates, e.g. a vector of auxiliary informations.

Due to the availability of large sets of informations about sample units and, in general, local areas of interest, provided by the full exploitation and analysis of the survey questionnaires, one of the most important questions is how to implement large sets of continuous and categorical variables in small area models based on composite estimates. In fact, many basic informations about units and areas are both continuous and categorical, and in many cases only the categorical ones can lead to appropriate assessments of specific issues. Another problem is that we assume in this linear models errors that have common variance.
In this paper we discuss about a general linear model of composite estimates, a model that consider the more general framework in which continuous and categorical predictors can help to evaluate appropriate small area estimates. A simulation is analyzed, based on the 1999 Italian Farm Structure Survey data.

References

Ghosh M., Rao J. N. K. (1994), *Small area estimation: an appraisal*, Statistical Science, Vol. 9, No. 1, pp. 55-93

Rao J. N. K. (2002), *Small area estimation with applications to Agriculture*, Proceedings of the Conference on agricultural and environmental statistical appications in Rome, Vol. III, pp. 555-564

Robinson G. K. (1991), *That BLUP is a good thing: the estimation of random effects*, Statistical Science, Vol. 6, No. 1, 15-51