A review of poverty mapping procedures

Isabel Molina
isabel.molina@uc3m.es
Universidad Carlos III de Madrid (Madrid, Spain)

J.N.K. Rao
jrao@math.carleton.ca
Carleton University (Ottawa, Canada)

Abstract

Efficient regional development policies require detailed assessment of the socio-economic development level of the regions within countries. Unfortunately, official surveys used to asses the living conditions of people have many times insufficient sample size to cover adequately all the regions within a country. Small area estimation techniques increase the effective sample size of the small areas by using auxiliary information through regression models. The most popular small area estimation approaches for poverty mapping will be revised, including several variants of the basic methods. Their properties, advantages and disadvantages will be discussed and open questions will be outlined. Results from an application with Spanish living conditions data will be shown.

Keywords: empirical Bayes; hierarchical Bayes; linear mixed models; poverty indicators; small area estimation.

1 Introduction

The socio-economic state of a population (let us say, a country) is usually assessed by governments using the results of surveys carried out by National Statistical Institutes. For this, the most reliable and detailed statistical measurement is required. Often, national surveys are not designed to give reliable statistical figures at local level. For instance, the Spanish Survey on Income and Living Conditions (SILC) is planned to produce direct poverty estimates at the Spanish Autonomous Communities, but it cannot not provide direct estimates for smaller subdivisions such as the Spanish provinces due to the small SILC sample sizes for some of these provinces. Here, a "direct estimate" for an area is an estimate that is calculated using only the available sample data from that area. These estimators are very inefficient for areas with small sample sizes. This problem has motivated the development of the scientific field called small area estimation (SAE), which finds more efficient "indirect" estimators for each area by "borrowing strength" from all the other ones. This can be done through the use of models that link all the areas, by assuming a constant relationship between the target variable and other auxiliary variables across areas, but at the same time allowing for additional area variation beyond that one explained by auxiliary variables.

Models used in this field can be classified into area level and unit level models. Area level models use only aggregated data over the areas. Aggregated data avoid confidentiality issues and are more readily available. However, the aggregation process often leads to some loss of information. A widely extended area level model is the Fay-Herriot (FH) model, introduced by Fay and Herriot (1979) to estimate mean per capita income in US small places.

In contrast, unit level models are established for the individual units. A popular model of this kind was proposed by Battesse, Harter and Fuller (1988) and will be called hereafter BHF model. It is a linear regression model that includes random area effects. These area effects represent the between area variation that is not explained by auxiliary variables. These models use more detailed information and often lead to greater gains in efficiency. Models with random effects belong to the general class of linear mixed models, extensively used in many other fields such as Biostatistics, Engineering, Econometrics and other Social Sciences.

In poverty mapping, the sample sizes of at least some of the target areas or domains are typically small and then small area estimation techniques are usually required. Besides, many poverty indicators are non linear functions of the target variable in the area units. For this reason, small area estimation techniques that deal with general non linear parameters have been developed.

To estimate poverty indicators in small areas, three popular approaches appear in the literature. The first one is not specially designed for non-linear parameters because it is based on the FH area level model, in which the dependent variable is the direct estimator of the target parameter in an area. This approach has been regularly used in the US Census Bureau, within the Small Area Income and Poverty Estimates (SAIPE) project (http://www.census.gov/hhes/www/saipe). This project produces estimates for states and counties of total persons in poverty by age groups, household median income and mean per capita income. These estimates are then used for the administration of federal programs and the allocation of federal funds to local jurisdictions. In Europe, the FH model was applied under the project EURAREA (http://www.statistics.gov.uk/eurarea) to estimate linear parameters such as mean income.

FH models require the specification of the sampling variances of the direct estimators. These variances have to be estimated but estimated values are typically used as if they were the true ones. However, due to the small area sample sizes, finding good estimators of these sampling variances is a problem. Another potential disadvantage of these models, apart from the fact of loosing information in the aggregation process, is that each particular poverty indicator has a different mathematical expression and therefore will require specific modelling.

Below we introduce two approaches based on unit level models. These models use much more detailed information and can be applied, once assumptions are fulfilled, automatically to general area parameters (not only poverty or inequality indicators) that are non linear functions of the target variable in the area units. A disadvantage of these methods is that they require great computational power, especially for large populations.

The first of these approaches is the ELL method due to Elbers, Lanjouw and Lanjouw (2003) and used by the World Bank. This method was especially designed to deal with complex non-linear poverty indicators. It assumes that the log incomes of the individuals in the population follow a unit level model similar to the BHF model, but including random effects for the sampling clusters instead of for the target areas. The method uses survey and census (or register) data that share the same auxiliary variables. After fitting the model to the survey data, ELL method generates by bootstrap a number of synthetic censuses from the fitted model using the census auxiliary data. The final estimates and their estimated variances are obtained by averaging over the bootstrapped synthetic censuses.

The second unit level approach, called empirical best/Bayes (EB) method, was recently introduced by Molina and Rao (2010) under the support of the European project SAMPLE. This method gives a Monte Carlo approximation of the estimator with minimum mean squared error or "best predictor". This is done under the assumption that there exists a transformation of the incomes of the individuals (or other welfare variable used to measure poverty), such that the transformed incomes follow the BHF model. Similarly as ELL method, it uses survey data in conjunction with census or register data. Mean squared errors of the EB estimators are estimated using a parametric bootstrap method. The EB method might provide estimators with notably better efficiency (approximately the "best") because it uses more extensively the sample information.

2 Extensions of basic unit level methods

Several variants of the basic EB method have appeared in the literature. This method requires to identify in the census or register, which are the survey units. Linking the census or register with the survey data might not be possible. The Census EB method has been proposed to avoid this linking step. Another variant was developed for estimation of computationally complex poverty indicators such as those that require sorting elements. In that case, computing EB estimates by the Monte Carlo approximation and using a bootstrap procedure for MSE estimation might become computationally unfeasible. For this reason, a fast EB method was proposed by Ferretti and Molina (2011). In this paper, the method was applied to estimate poverty fuzzy monetary and supplementary indicators (Betti et al., 2006). A very recent alternative to the EB method is the hierarchical Bayes (HB) procedure introduced by Molina, Nandram and Rao (2013). This method reduces a lot the computational time because it avoids the bootstrap procedure for MSE estimation. This is possible thanks to the fact that HB methods provide the whole picture of the posterior distribution of the target parameters. Posterior variances are used as uncertainty measures of the resulting HB estimators. Another advantage of the HB procedure is that any other summary of the posterior distribution, such as credible intervals, can be easily obtained.

Concerning ELL method, lately the method is being applied using a model with random domain effects instead of cluster effects, which makes it more comparable with the EB method. Parametric bootstrap with re-estimation of the model parameters similar to the bootstrap procedure applied by Molina and Rao (2010) is also being applied for MSE estimation of the ELL estimator. Two-fold models with random cluster effects nested within the area effects are being currently studied for both EB and ELL methods. Models that assume mixtures of normal distributions for random area effects and individual model errors are also being recently proposed. Robust procedures based on M-quantiles are also being proposed. All of these methods will be reviewed, mentioning advantages and disadvantages of each one. Finally, fields of improvement and potential future extensions will be outlined.

References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association* **83**, 28–36.

Betti G., Cheli B., Lemmi A. and Verma V. (2006). Multidimensional and Longitudinal Poverty: an Integrated Fuzzy Approach, in Lemmi A., Betti G. (eds.) *Fuzzy Set Approach to Multidimensional Poverty Measurement*, 111–137, Springer, New York.

Elbers C., Lanjouw J.O. and Lanjouw P. (2003). Micro-level estimation of poverty and inequality. *Econometrica* **71**, 455–364.

Ferretti, C. and Molina, I. (2011). Fast EB method for estimation complex poverty indicators in large populations. *Journal of the Indian Society of Agricultural Statistics* **66**, 105–120.

Molina, I. and Rao, J.N.K. (2002). Small area estimation of poverty indicators. *Canadian Journal of Statistics* **38**, 369–385.

Molina, I., Nandram, B. and Rao, J.N.K. (2013). Small area estimation of general parameters with application to poverty indicators: a hierarchical Bayes approach. Unpublished manuscript.