Small area estimation of poverty indicators using interval censored income data

Extreme poverty rates have been cut by more than half since 1990. While this is a remarkable achievement, it is still one of the main goals defined by the United Nations to eradicate extreme poverty by 2030. To fight poverty, it is essential to have knowledge about its spatial distribution. Small area methods enable the estimation of poverty indicators at a geographical level where direct estimation is either not possible or very imprecise, due to a lack of sample size.

Among a variety of small area estimation methods (SAE), one popular approach for the estimation of poverty indicators is the empirical best predictor (EBP) proposed by Molina and Rao (2010). However, direct parameter estimation is not possible, when the dependent variable of the underlying nested error regression model, such as income or consumption, is censored to specific intervals, due to confidentiality constraints or other reasons.

Therefore, this work introduces two estimation methods, which enable the estimation of the regression parameters of the nested error regression model under these circumstances. The proposed estimation methods are based on the expectation maximization (EM) algorithm and on the stochastic expectation maximization (SEM) algorithm. Since the EBP approach strongly relies on Gaussian assumptions of the error terms, data-driven transformations are incorporated into the algorithms to handle departures from normality. The estimation of the mean squared error (MSE) of the EBP is facilitated by a parametric bootstrap which captures the additional uncertainty due to the interval censored dependent variable.

Empirical evaluations indicate that the EM- and SEM-algorithm are able to handle interval censored data and outperform naive estimation procedures, like the EBP based on regression on the midpoints of the intervals, or direct estimation. In addition, we evaluate the performance of the data-driven transformations in this context and assess the proposed parametric bootstrap for MSE estimation. Finally, we illustrate the methodology using data from the German micro-census. The micro-census is a 1% sample of German households where income is censored to intervals due to confidentiality constraints.