

SYNTHETIC DATA SOURCES IN THE SPATIAL ANALYSIS OF POVERTY IN POLAND

Wojciech Roszka

Department of Statistics, Poznan University of Economics, POLAND, wojciech.roszka@ue.poznan.pl

One of the key elements of the state's social policy is the poverty prevention. Information about its spatial diversity is very helpful in the context of allocation of resources and to take measures to prevent its growth. Characteristics linked to the poverty estimates are mainly based on European Union Statistics on Income and Living Conditions (EU - SILC), which is a sample survey conducted annually in all countries of the European Union. Limitations associated with the sample size mean that the at-risk-of-poverty rate are estimated at most to the level of the provinces (NUTS 2, voivodeships in Poland).

The application of the small area estimation methods (SAE) allows better quality estimation without increasing the sample. The problem in the context of poverty mapping in Poland was, among others, raised by Wawrowski (2014), who used a Fay - Herriot model to estimate the at-risk-of-poverty rate at NUTS 3 level. Simultaneously with the model approach methods, techniques for creating synthetic full-coverage datasets on the basis of available information are being developed (Haslett et al. 2010). These techniques are already of interest to the European Union in the context of the spatial diversity of the living conditions of the population (Alfonso et al. 2011).

The synthetic datasets are being created using dynamically developing methods of spatial microsimulation. They allow to create multi-dimensional estimates for small domains. Especially synthetic reconstruction methods will be presented. With the joint usage of statistical matching methods (Raessler 2002) and iterative proportional fitting models (Rahman 2008) it is possible to construct the synthetic micro-populations at a small area level in such a way that all known constraints at the small area level are reproduced.

The aim of the paper is to assess the possibility of obtaining multivariate estimates of acceptable quality for at-risk-of-poverty at NUTS 4 (poviats in Poland) using a synthetic data set created on the basis of EU - SILC 2011 dataset and information from the census. The resulting estimates will be evaluated by their reliability, consistency and quality, as well as a comparative analysis will be carried out with the results obtained by "classical" SAE models.

References

- Alfonso A., Filzmoser P., Hulliger B., Kolb J-P., Kraft S., Munnich R., Templ M., 2011, *Synthetic Data Generation of SILC Data*, European Commission, Community Research, AMELI Project
- Haslett S., Jones G., Noble A., Ballas D., 2010, *More or Less? Comparing small area estimation, spatial microsimulation, and mass imputation*, Section on Survey Research Methods JSM, American Statistical Association
- Raessler S., 2002, *Statistical Matching. A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*, Springer, New York, USA
- Rahman A., 2008, *A Review of Small Area Estimation Problems and Methodological Developments*, Discussion paper 66, NATSEM, University of Canberra
- Wawrowski ., 2014, *Wykorzystanie metod statystyki małych obszarów do tworzenia map ubóstwa w Polsce [The use of small area estimation methods for poverty mapping in Poland]*, Wiadomości Statystyczne 9/2014, Polskie Towarzystwo Statystyczne [*Polish Statistical Association*]