Name, E-Mail: Tim Goedemé; tim.goedeme@ua.ac.be

Institution: Herman Deleeck Centre for Social Policy (University of Antwerp); Belgium

Titel: How much confidence can we have in EU-SILC? Complex sample designs and the standard error of the Europe 2020 poverty indicators

Abstract: Currently, the Community Statistics on Income and Living Conditions (EU-SILC) is the single most important data source for cross-national comparative research on income and living conditions in the European Union. As EU-SILC consists of a sample of European households, point estimates should be accompanied by appropriate standard errors and confidence intervals. This is especially so if indicators are constructed for measuring progress towards pre-defined targets such as those of the Europe 2020 strategy. All too often this has been neglected in European poverty research and official publications. In contrast, this paper pays explicit attention to the calculation of standard errors and confidence intervals of EU-SILC based indicators and enquires which is the best way forward for EU-SILC users when estimating the sampling variance.

Apart from imputation and weighting, the sample design strongly affects the standard error of estimates. Therefore, accurate information on the sample design is crucial, especially for a database like the EU-SILC User Database (UDB) which contains data on 27 European countries which employ many different complex sample designs. However, for many countries important aspects of the sample design are not very well documented. Moreover, crucial information on the sample design is incomplete in the EU-SILC User Database and there are several options for handling this lack of information.

Therefore, in this paper, I bring together for all countries included in the EU-SILC 2008 UDB indispensible information on the sample design. Additionally, I compare actual sample designs with the available information in the EU-SILC UDB and in the complete dataset available to Eurostat. Furthermore, on the basis of Eurostat data with complete information on the sample design I explore which variables are best used when analysing EU-SILC for adequately computing standard errors, as there are several options for treating the lack of information in the EU-SILC UDB. I illustrate the sampling variance and the importance of various assumptions with regard to the sample design by presenting results for the official Europe 2020 poverty indicators. It is shown that neglecting the sample design can lead to a serious under-estimation of the standard errors. In addition, it is discussed how researchers using the EU-SILC UDB could best take account of the sample design for appropriately estimating standard errors.