

Statistical matching of EQLS and EU-SILC: A case study on public services

Daphne Ahrendt, Eurofound, Dublin. dap@eurofound.europa.eu
Tadas Leoncikas, Eurofound, Dublin. tle@eurofound.europa.eu
Irene Riobóo, Rey Juan Carlos University, Madrid. irene.rioboo@urjc.es

Abstract

The increasing growth of collected data together with the continuous research interest on the interrelationships of several socio-economic aspects, calls for robust methodologies to combine different sources of information. In this line, the current study examines the benefits of the statistical matching between the European Quality of Life Survey (EQLS) and the EU Statistics on Income and Living Conditions (EU-SILC) in relation to their new modules on public services included in 2016.

Both sources play a key role in the analysis of living conditions and quality of life. Therefore, the possibility to combine the different angles covered by the EQLS and the EU-SILC makes their statistical matching an extremely interesting approach.

The literature shows that some exercises to link them were carried out in the past, concluding that the difficulties found could not guarantee the quality of the integration. Nevertheless, both the efforts made in standardize the core variables and the inclusion of the new modules on public services make a new matching attempt worthwhile.

Considering that those modules provide different and complementary information on access, use and quality of the services, the possibility to merge them creating a synthetic dataset would give a broader picture, expanding the focus of both sources and incorporating different angles without increasing response burden or costs.

The study started with the harmonization of both sources. The comparison of their methodologies and metadata implied two relevant adjustments in relation to the reference population and the country coverage. On the one hand, EU-SILC had to be adjusted to persons aged 18 and over due to the EQLS reference population is all persons aged 18 or older who reside in the country at the time of data collection. On the other hand, EQLS had to be adjusted to the countries available for the EU-SILC when the study was carried out, meaning that the study covers 25 countries: the UE Member States except Ireland, Italy, Luxembourg and Malta, plus Serbia.

Next steps implemented were in line with the classical methodological guidelines used in statistical matching:

- Identify common variables, exploring the correspondences that the EQLS's questions have in the EU-SILC, analysing their conceptual and statistical consistency.

- Determine target variables, selecting the set of non-overlapping variables of interest (available only in one of the sources) that will be matched. Those variables have to be related to the research questions, hence, in the framework of this study, they have to be focused on some aspects of public services.
- Select matching variables, meaning those common variables high correlated with the target variables and with a high explanatory power in the original sources. They are the variables to be used to link both sources for predicting the target variables in the synthetic dataset.
- Implement statistical matching, using different models to match the EQLS and the EU-SILC target variables through the matching variables and creating in that way a synthetic dataset that combine both sources. The software used has been R and particularly its package StatMatch.
- Evaluate results, assessing the quality of the results provided by each model implemented and identifying the best alternative.

The main conclusion obtained in this study is that some of the models used in the matching procedure are satisfactory and provide synthetic datasets that can be used to carry out further analysis in the field public services.

Keywords: statistical matching, combined analysis, StatMatch, EQLS, EU-SILC, public services.