The Survey Statistician, 2021, Vol. 84, 18–23.

Reducing non-sampling errors to improve the measurement of poverty and social exclusion in Europe

Peter Lynn

University of Essex, UK, plynn@essex.ac.uk

Abstract

Non-sampling errors can seriously influence statistical estimates based on survey data. A recently-completed initiative has attempted to map out the influence of non-sampling errors, and to identify ways in which they could be reduced, on EU-SILC. This article provides a overview of the work that has been carried out.

1 Introduction

The European Union Statistics on Income and Living Conditions (EU-SILC) instrument is the reference source for comparative statistics on income, wellbeing and social inclusion in the EU. EU-SILC takes the form of a survey carried out in all EU member states, with both cross-sectional and longitudinal components, though some data items are collected directly from registers in some countries. Data are collected from over a quarter of a million households across Europe each year. For the longitudinal component, households are asked to participate in each of four consecutive annual waves.

The third Network of EU-SILC Researchers (Net-SILC3) was funded by Eurostat to carry out a programme of research activities and to disseminate best practice to National Statistical Institutes (NSIs) over a 4-year period (2016 to 2020). One of the two clusters of Net-SILC3 activities focussed on sources of non-sampling errors in EU-SILC. This article focuses on the activities of that cluster.

Non-sampling errors are systematic or variable errors that arise from aspects of the survey process other than sample selection. These include errors due to frame under-coverage, non-response, field work, measurement and processing. Such errors are important as studies in various contexts have shown them to often be at least as influential as sampling error, and sometimes more so. They can, however, be reduced through a combination of improved data collection practices and post-survey adjustment methods, but this requires understanding of the nature of the sources of error. The Net-SILC3 cluster of 11 work packages was designed to identify the main sources of non-sampling errors in EU-SILC, to describe the nature and impact of each type of error, and to produce guidance on reducing them. Ten of the work packages were to deliver original research on various aspects of non-sampling errors, while the eleventh was to organise a series of best practice workshops. The network involved six NSIs (Austria, Finland, Latvia, Netherlands, Serbia, UK) and six academic and independent partner organisations (University of Essex, University of Antwerp, University of Manchester, GESIS, Sciensano, Lyberg Survey Quality Management). Additionally, two other NSIs (Sweden and Slovenia) participated voluntarily in one of the work packages.

The remainder of this article summarises some of the main findings and conclusions arising from the ten work packages. It should be noted that the conclusions are those of the authors of the research and do not necessarily reflect the views of Eurostat. Further details of the research can be found in the forthcoming book arising from the project, the contents of which are summarised in Box A. Chapters 3 through to 26 of the book report the work carried out by the members of Net-SILC3, while

the last three chapters are invited contributions from researchers working on other international surveys whose research has implications that are transferable to EU-SILC. The following summary indicates the chapter number for each study mentioned, allowing the reader to identify the author of the research by referring to Box A.

2 Population Coverage

EU-SILC covers only the private household population. Those living outside of private households consist mainly of residents of institutions of various kinds. The size of the institutionalised population varies between countries but is generally below 2.5% (amongst EU countries, this proportion is exceeded only in Slovakia and Sweden). Their characteristics also vary: they are disproportionately male in Greece and Lithuania but female in Germany and Belgium. In most countries, the institutionalised population is older and less educated than the household population. A case-study in Finland found the non-household population to have substantially lower levels of income than the household population. If viewed as estimates of the characteristics of the total population, EU-SILC estimates may therefore suffer from non-coverage bias (Chapter 3).

Within the private household population too there may be under-coverage if some households are missing from the sampling frames. It turns out that many of the sampling frames used for EU-SILC have high levels of coverage. What little under-coverage of this kind can be detected tends, in register countries, to be related to delays in the registration of recent immigrants. However, there is a severe paucity of information about under-coverage in many countries and research in this area could be warranted (Chapter 4).

3 Survey Nonresponse

EU-SILC survey response rates vary considerably between countries, and between years within countries, from less than 50% in some cases, to over 90% in others. Mean response rate has declined from 70% in 2006 to 60% in 2017. Consequently, there is scope for nonresponse bias to affect survey findings, and particularly comparisons between countries.

The composition of EU-SILC responding samples was compared to comparable Census findings (Chapter 5). The extent to which the sample appeared representative of the Census population varied greatly between countries and seemed only weakly related to response rate. Samples were generally less representative in terms of economic activity, education level and citizenship than in terms of age and sex. But again there were big differences between countries with high levels of education-related nonresponse bias in Belgium and the UK, but very low levels in Italy and the Netherlands.

Panel attrition across the waves of data collection was found to be associated with survey variables related to income, health and wellbeing, but the associations varied greatly between countries. However, the magnitude of the resultant nonresponse bias was generally rather small (Chapter 6).

Allowing survey data to be collected from a proxy respondent can reduce the level of nonresponse, but it can involve a trade-off between nonresponse error and measurement error. For EU-SILC, proxy response is supposed to be used only in exceptional circumstances but it was found that one in five interviews were in fact carried out by proxy, with prevalence varying greatly between countries, from below 6% in Greece, Slovakia, Switzerland and Sweden to over 40% in Denmark, Finland and Croatia (Chapter 7). There are several countries in which the extent of proxy responding has been steadily increasing over the years, while in others it has decreased. Over the waves of a panel the extent of proxy responding tends to increase, while in some countries the level seems to peak at wave 2. Proxy reporting is particularly common for young adults (aged 16 to 24), students and those undertaking military service.

The extent to which best practice in procedures for limiting nonresponse (bias) is implemented varies between countries. An example is interviewer training in nonresponse avoidance skills. Some good

practices, such as the use of targeted nonresponse procedures, drawing upon information from the sampling frame or from wave 1, seem hardly to be used at all (Chapter 8).

4 Weighting and Imputation

Statistical adjustment through weighting can reduce the impact of both nonresponse bias and sampling variance. For a comparative survey, it may be particularly important that this is done in a comparable way in each country. For EU-SILC, responsibility for the production of weights lies with the national teams, but guidance is provided on the approaches and methods that should be used. In general – but not universally – EU-SILC weighting was found to be done according to the guidance. But practices still vary considerably (Chapter 9).

A study of data from five countries - Finland, Latvia, The Netherlands, Slovenia, and Sweden – examined whether the addition of registered income data improves the performance of the weight calibration (Chapter 10). The gain in precision varies was found to vary between countries (being particularly high in the Netherlands) and between estimates (being better for estimates of poverty or social exclusion than for employment-related estimates).

Two approaches to weighting in the context of a modular design were compared: composite and two-phase calibration (Chapter 11). Modular design refers to describe surveys which have a common (or core) set of questions administered in the same way, plus multiple modules of questions each administered to separate sub-sample. UK data, in which one of the modules contains the EU-SILC questions, were used in the comparison. It was found that either approach improved precision, compared to ignoring the modularity and treating the EU-SILC module as a stand-alone survey. The composite method may be preferable if the design has only two modules.

The longitudinal component of EU-SILC effectively results in a rotating panel design, with a new 4-year panel starting each year. There are several possible ways to adjust for attrition in the production of cross-sectional weights for a rotating panel design. Four such methods were compared for 27 countries (Chapter 12). It was concluded that the choice of predictor variables in nonresponse models may be more important than the choice of modelling method. All of the methods performed considerably better for estimating low income than for estimating poor health. This was most likely a result of the predictors chosen for the models being more pertinent to income than to health.

While EU-SILC weighting practices are generally good, best practice is not always used (Chapter 13). Furthermore, there is considerable variation in approaches between countries, particularly in the choice of auxiliary variables and the assumptions made about eligibility. It is suggested that EU-SILC might benefit from recommending a specific (standardised) approach to be used for imputation of eligibility status, for nonresponse modelling, and for combining multiple panels in nonresponse adjustment – rather than allowing countries to choose an approach.

Levels of item nonresponse to income questions on EU-SILC are generally low, and are extremely low (sometimes even zero) in countries that collect this information from registers rather than through survey interviews. However, practices vary in terms of how countries use flag variables to indicate the nature of missing data and in terms of the information published regarding item nonresponse in quality reports. This makes it hard to make comparisons (Chapter 14).

Imputation methods used to deal with item nonresponse vary between countries (Chapter 15), with many different kinds of methods in use. Four countries do no imputation as all income values are taken from official registers. Surprisingly, several countries use mean or median imputation. Others use regression methods or donor methods such as hot deck or nearest-neighbour. Only three countries use some kind of repeated imputation (Switzerland, Italy and Croatia). The effect of imputation on estimates was found to differ between countries. A simulation study showed broadly similar estimates arising from a number of different imputation approaches. Documentation of the

imputation methods used was found to need improvement. Some degree of harmonisation of the strategies used could also be beneficial.

Most countries use only cross-sectional imputation methods, even for the EU-SILC longitudinal data. The advantages of longitudinal methods such as as 'Last value carried forward' - with or without uprating - or the row and column method of Little and Su (1989) would seem to be worth exploring. A simulation study showed that longitudinal imputation performed well, though effects on estimates were small in magnitude (Chapter 17).

For EU-SILC, some countries collect gross income, some net income, and some both. This results in a need to convert reported amounts from net to gross or vice versa, in order to be able to carry out analysis on a consistent basis and make comparisons. This can be done in a deterministic way if sufficient knowledge is available regarding the country's systems of taxes and social insurance contributions and if sufficiently detailed data has been collected to establish for each household and individual the taxes and contributions that should apply. This is not always, or often, the case, so stochastic statistical methods are often used instead. These methods vary between countries, as do the proportion of cases for which net-gross conversion is needed (Chapter 16). Some countries use micro-simulation methods but none use EUROMOD, the tax-benefit microsimulation model for the European Union and UK. Use of EUROMOD would aid consistency of approach between countries.

5 Comparability and Validity of Measures

The form of survey questions can also affect measurement, and hence comparability between countries. Practices regarding how EU-SILC income questions are asked, the examples that are given, and the level of detail (and number of questions) with which the information for target variables is collected vary greatly between countries (Chapter 18). As a result, variables may not be comparable. Additionally, some countries have moved from using survey measures to register data without evaluating the impact on longitudinal measures or time series.

Some questions on health and health care too are asked in ways which vary between countries. For example, the number of visits to a health care provider is asked as an open or closed question, with or without filters. Questions about basic functions may or may not specify that difficulties walking or climbing steps should be assessed without the use of any assistance or device. But body mass index and the consumption of fruit and vegetables are measured in an almost identical way in all countries (Chapter 19).

With regard to housing measures (Chapter 20), some difficulties were identified with the measure of number of rooms. Explanations of which types of rooms to include, and of how to treat shared rooms, vary between countries and there is no explicit guidance on how to treat open-plan spaces. This measure is important as the official indicator of over-crowding is based upon it. The details collected about mortgages also vary, including whether or not interest payments are collected separately from capital payments. In principle, this distinction is important as interest is a form of expenditure while capital payments constitute savings.

The production of goods for a household's own consumption (mainly food, but also some other things such as wood for fuel) is in principle an income and can be an important component of poverty alleviation. Some countries attempt to measure this for EU-SILC while others do not. Those that do, differ in how the income is measured (monetised): some ask households to assess the total value while others (Cyprus, Spain, Czechia, Slovakia) record the quantity of goods produced in each of several categories and subsequently convert this at market prices (Chapter 21).

The percentage of households reporting own consumption goods exceeds 20% in ten countries but the percentage of total income accounted for by own consumption is much smaller, exceeding 1% in only four countries (Latvia, Croatia, Estonia, Portugal). The inclusion of own consumption goods in household income makes very little difference to key estimates, even for subgroups such as rural

populations and households in the lowest income quintile. There is not therefore a strong case for amending the EU-SILC measure of total disposable income to include own consumption goods (Chapter 22).

6 Survey modes and Survey Processing

The data collection modes, and mixes of modes, used for EU-SILC vary considerably between countries and have changed over time within countries. Modes are known to affect both participation and measurement, so this may be a concern for comparability. Various strategies can be used to prevent (minimise) mode effects and to adjust for them. There are a number of steps that could be taken to enhance knowledge of mode issues amongst NSI teams responsible for EU-SILC, to reduce differences in approaches and to facilitate adjustment (Chapter 23).

Considerable variation between countries was identified in the data collection modes used, interviewer workloads, interviewer training, field quality control procedures, coding systems and editing systems (Chapter 24). Greater standardisation of interviewer training and fieldwork quality control procedures is suggested along with co-ordination and co-operation in a proactive move towards more widespread adoption of web data collection (in 2018 only 6 of 30 countries used web data collection, with those 6 using it in different ways and at different stages).

A series of studies in the UK tested whether a general population sample would engage with an online survey of finances and whether design features such as telling them in advance that they would need to refer to financial documentation are helpful. Results were broadly positive and can help inform a move to online data collection for EU-SILC (Chapter 25).

A cost-benefit analysis using Dutch EU-SILC data (web and CATI) suggests that the use of a re-interview design to produce estimates of mode-specific selection and mode-specific measurement biases may improve accuracy if estimates are subsequently adjusted for measurement biases. The study should be replicated for other countries and designs that include face-to-face (Chapter 26).

Reference

Little, R. J. A. & Su, H.-L. (1989), Item Non-Response in Panel Surveys. In D. Kasprzyk, G. Duncan and M. P. Singh (Eds.), *Panel Surveys*. New York: John Wiley. pp. 401-425.

Box A: Contents of Forthcoming Book

Title: Improving the measurement of poverty and social exclusion in Europe: reducing non-sampling errors
Editors: Peter Lynn and Lars Lyberg
Publisher: Publications Office of the European Union
Contents:

Introduction

1. *Handling non-sampling errors in comparative surveys*, Lars Lyberg (Demoskop Inc, Sweden)
2. *Investing in statistics: EU-SILC*, Emilio Di Meglio, Didier Dupré and Sigita Grundiza (Eurostat)

Population coverage and survey nonresponse

3. *The effect on EU-SILC of exclusions from the target population*, Tara Junes (Statistics Finland)
4. *Frame errors in the EU-SILC: Under-coverage*, Tara Junes (Statistics Finland)
5. *Representativeness of first wave EU-SILC response and response rates over time*, Natalie Shlomo (University of Manchester, UK), Annemieke Luiten, (Statistics Netherlands) and Barry Schouten (Statistics Netherlands and Utrecht University)
6. *Impact on representativeness of EU-SILC panel attrition*, Barry Schouten (Statistics Netherlands and Utrecht University) and Annemieke Luiten, (Statistics Netherlands)
7. *The effect of proxy responses on non-response error*, Peter Lynn (University of Essex, UK)
8. *Current best practice in minimising nonresponse error in panel surveys*, Peter Lynn (University of Essex, UK)

Statistical adjustment: weighting and imputation

9. *Review of EU-SILC weighting methods*, Mārtiņš Liberts (CBS Latvia)
10. *Use of registers in calibration*, Mārtiņš Liberts (CBS Latvia)
11. *Weighting for a modular structure*, Andy Fallows (Office for National Statistics, UK)
12. *Weighting together panels for cross-sectional estimation*, Olena Kaminska (ISER, University of Essex, UK)
13. *Current best practice in weighting for a rotating panel*, Gareth James (Office for National Statistics, UK), Mārtiņš Liberts (CBS Latvia) and Peter Lynn (ISER, University of Essex, UK)
14. *Item non-response in EU-SILC income variables: levels, characteristics and documentation*, Richard Heuberger (Statistics Austria)
15. *Imputation for income variables in EU-SILC*, Sophie Psihoda (Statistics Austria), Nadja Lendle (Institute of Applied Statistics, JKU Linz), Richard Heuberger (Statistics Austria) and Thomas Glaser (Statistics Austria)
16. *Methods for net-gross conversion in EU-SILC*, Richard Heuberger (Statistics Austria)
17. *Longitudinal imputation of EU-SILC income variables*, Nadja Lendle (Institute of Applied Statistics, JKU Linz, Austria) and Matthias Till (Statistics Austria)

Comparability and validity of measures

18. *Lessons and recommendations regarding the comparability of the EU-SILC income variables*, Tim Goedemé (University of Antwerp, Belgium and University of Oxford, UK) and Lorena Zardo Trindade (University of Antwerp, Belgium)
19. *The validity and comparability of EU-SILC health variables*, Stefaan Demarest (Sciensano, Belgium) and Rana Charafeddine (Sciensano, Belgium)
20. *Comparability and validity of the EU-SILC housing variables*, Ross Bowen, Callum Clark and Paola Serafino (Office for National Statistics, UK)
21. *Methods for collecting data on production for own consumption*, Tijana Čomić (Institute of Economic Sciences, Serbia)
22. *The impact of own consumption on income distributions and key EU income-based indicators*, Tijana Čomić (Institute of Economic Sciences, Serbia)

Survey modes and survey processing

23. *Mode issues in comparative surveys*, Lars Lyberg (Demoskop Inc, Sweden), Peter Lynn (University of Essex, UK) and Barry Schouten (Statistics Netherlands and Utrecht University)
24. *Preventing and mitigating effects on data quality generated by mode of data collection, coding and editing*, Sophie Psihoda (Statistics Austria), Nadja Lamei (Statistics Austria) and Lars Lyberg (Demoskop Inc, Sweden)
25. *The potential role of EU-SILC as part of an integrated social survey: The case of ONS*, Ria Sanderson and Pete Betts (Office for National Statistics, UK)
26. *A cost-benefit analysis of EU-SILC mode effect decomposition: Three case studies*, Barry Schouten (Statistics Netherlands and Utrecht University)
27. Mode and Web Panel Experiments in the European Social Survey: lessons for EU-SILC, Rory Fitzgerald and Eva Aizpurua (City, University of London, UK)
28. *Interviewers and their impact on survey quality: lessons for EU-SILC from the European Social Survey*, Geert Loosveldt (KU Leuven, Belgium)
29. *The importance of occupation coding quality: lessons for EU-SILC from SHARE and other international surveys*, Kea Tijdens (University of Amsterdam, Netherlands)