A new database on the contents and comparability of the income variables in EU-SILC: MetaSILC 2015

Lorena Zardo Trindade[1]*, Tim Goedemé[1,2]*

[1]Herman Deleeck Centre for Social Policy, University of Antwerp, Antwerp, Belgium; [2]Institute for New Economic Thinking at the Oxford Martin School and Department of Social Policy and Intervention, University of Oxford, Oxford, UK

Abstract The MetaSILC 2015 database is a new addition to the toolbox of EU-SILC users and producers. It documents how individual income components are aggregated into the EU-SILC target variables. Even though general and country-specific descriptions of income target variables are available in the EU-SILC methodological guidelines and in the national quality reports, it is often not clear how exactly each of the national income components is classified and aggregated into a target variable in practice. On the basis of a survey among national statistical institutes, we compiled a database which maps the exact classification of income components onto the EU-SILC target variables. The focus of the database is on EU-SILC 2015, covering 26 EU-SILC countries. The database contains information on the composition of variables regarding total income before and after transfers; income from benefits, work and capital; social contributions and taxes; as well as on recent and planned changes to the composition of the variables.
JEL classification: C8, D1, D3

*For correspondence: Lorena.ZardoTrindade@uantwerpen.be; tim.goedeme@spi.ox.ac.uk

1. Introduction

The EU Statistics on Income and Living Conditions (EU-SILC) is one of the principal datasets for studying income and living conditions in Europe. The data are widely used for studying poverty and inequality in the EU and serve as the data source for official poverty reduction targets. Furthermore, EU-SILC is often used for monitoring the social situation both within countries and across countries, and for carrying out both ex-ante and ex-post policy evaluations. It is also the main dataset for European countries covered by EUROMOD, the tax-benefit microsimulation model for the European Union. The quality and comparability of the income variables in EU-SILC are therefore of paramount importance. The data collection of EU-SILC follows a so-called ex-ante output harmonization model: target variables that countries must collect are specified, but the exact way in which these variables are collected is up to the participating countries. Furthermore, the fact that tax-benefit systems differ considerably across countries, makes it sometimes difficult to aggregate income components in a consistent way into the relatively generic target variables of EU-SILC. As a result, the collection and categorisation of specific income components varies considerably across countries.

Although general and country-specific descriptions of income target variables are available in the EU-SILC methodological guidelines and in the national quality reports, it is often not clear how exactly each of the national income components is classified and aggregated into a target variable. This implies that researchers sometimes do not realise the wide variety in types of income sources that some variables cover, potentially undermining the validity of the conclusions of their analysis. Also, this constitutes an important hurdle for researchers who want to use EU-SILC for tax-benefit

microsimulations, given that it is essential to know which income sources are categorised under which variable in order to simulate tax liabilities and social security contributions, or to validate simulated benefits and allowances. Often, it can be a cumbersome process to identify the exact classification of each income component.

Therefore, in the context of Net-SILC 3[1], we created the MetaSILC 2015 database that documents how individual income components are aggregated into the EU-SILC target variables (*Goedemé and Zardo Trindade, 2020a*). MetaSILC 2015 allowed us to produce a report with a detailed analysis of compliance with the Eurostat guidelines and definitions, looking into inconsistent classifications of income components across countries and changes to the variables across time (*Goedemé and Zardo Trindade, 2020b*). Additionally, in a paper published by Eurostat we formulated recommendations about how to improve the quality and comparability of the income variables in EU-SILC (*Zardo Trindade and Goedemé, 2020*). In what follows we present MetaSILC 2015 in some more detail and discuss the main outputs as well as future updates.

2. The database

The focus of MetaSILC 2015 is on the 2015 wave of EU-SILC, and covers 26 EU-SILC countries. The database contains information on the composition of variables regarding total income before and after transfers; and the more disaggregated variables on income from benefits, work and capital; social contributions and taxes (see Table A.1). It contains additional information on the measurement of income in kind associated with a company car, imputed rent and the measurement of self-employment income. It also documents changes to variables in the period 2010–2015 and prospective changes in subsequent years and contains more information on outlier detection and treatment. The database is freely available in excel to EU-SILC users and other interested parties (*Goedemé and Zardo Trindade, 2020a*).

The main source of the database consists of two questionnaires that were sent out to National Statistical Institutes (NSIs) between July 2016 and January 2017. The NSIs of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, France, Germany, Greece, Hungary, Italy, Latvia, Luxembourg, Malta, the Netherlands, Poland, Republic of Serbia, Slovakia, Slovenia, Spain, Sweden and United Kingdom responded to both questionnaires, while the NSIs of Finland and Portugal provided partial information[2]. Responding NSIs were asked to list all income sources covered in EU-SILC, at the most detailed level, and how they were classified into the EU-SILC target variables. With regard to total income before and after transfers and income from benefits, they were asked to report if the equation used to compute all four variables (HY010, HY020, HY022 and HY023) was in accordance with the Eurostat guidelines (*Eurostat, 2016*). In the case they were not, they were asked to provide additional information on the deviations. With respect to income from benefits, income from work and from other sources, social contributions and taxes, respondents were asked to provide more details about the composition of each target variable. For each of the income components used to compute the target variables, we collected information on the official name (national language) and code in the national EU-SILC survey; the equivalent name in English; the target variable code and name; the question in the EU-SILC questionnaire, the source of the income information used; the level of aggregation when it was collected and information on whether it was collected gross and/or net.

Complementary information to reported income components was collected over the course of 2017, 2018 and 2019. In addition to the information from the survey among national statistical institutes, the accompanying report was supplemented with information from the national quality reports, the comparative quality reports, the national questionnaires and other sources. [3] NSIs were asked to answer follow-up questions for clarification and to review a preliminary version of the database and report. All comments received were included in the report and in the dataset when appropriate.

---

1. The Third Network for the Analysis of EU-SILC, funded by Eurostat, see https://ec.europa.eu/eurostat/cros/content/third-eu-silc-network-income-and-living-conditions-net-silc3_en.
2. Ireland, Lithuania, Romania, Norway, Iceland, Switzerland did not take part even though they have data available for *Eurostat (2016)*
3. Some of the information was also cross-checked with information from the European System of Integrated Social Protection Statistics (ESSPROS), the Mutual Information System on Social Protection (MISSOC,) as well as the EUROMOD country reports.

3. The various outputs

Using the MetaSILC 2015 database, a report (*Goedemé and Zardo Trindade, 2020b*) and a working paper (*Zardo Trindade and Goedemé, 2020*) were compiled. In the report, for each income variable used to construct total disposable household income (see Table A.1), we discuss compliance with Eurostat guidelines, misclassifications and omitted income sources that could undermine cross-national comparability. The sections for each target variable contain a summary with main findings, the variable definition in the Eurostat guidelines, general results on cross-national comparability, and a section with detailed remarks by country. In the working paper, we bring together the main findings from the report, highlight prospective changes to EU-SILC in response to our report, and formulate some recommendations regarding how comparability can be improved in the future. Apart from updating and expanding the MetaSILC 2015 database, our main recommendations include the establishment of an expert panel which would help Eurostat to improve the description of the target variables and help NSIs to aggregate all income sources into the EU-SILC target variables in a way that is consistent across countries. Also, the ex-ante output harmonization model could be strengthened by providing concrete examples of a 'preferred way' to ask questions regarding income in the EU-SILC question-naires, especially with regard to the level of detail with which questions are asked and the examples that are given as income sources to be included (eg, when asking about income from capital), in line with what is already done for the variables on material deprivation. Other recommendations include setting up a repository for research that documents the impact on the income variables of moving from survey data to register data, and creating more consistency, and transparency, regarding the way in which NSIs convert net incomes to gross incomes and the other way around.

4. Future updates

It is hard to find detailed information on the exact implementation of EU-SILC in each country, and to document the comparability of the income variables. In our opinion, the MetaSILC 2015 database, set out as a onetime survey among the data producers, proved to be very helpful for collecting detailed metadata on how specific income components were aggregated into the EU-SILC income target vari-ables. Therefore, we believe it is a useful strategy to repeat and expand MetaSILC in the future, pref-erably by Eurostat, as it clearly provides essential information that is nowhere else available. Ideally, this would be done during the data production process, as at that point all required information is already available to NSIs. We are confident that for future updates, it should be possible to limit the required time investment on the part of NSIs. One can also think about ways to facilitate data users to share more information on the comparability of specific variables for specific research purposes.

A more ambitious expansion that could be considered, and would be very useful to many users, is to complete the MetaSILC dataset with an institutional description of each income component (espe-cially benefits) covered by the target variables. This would allow individual users or ex-post harmon-isers like Eurostat and the Luxembourg Income Study (LIS) to save considerable time when looking for institutional information before creating aggregates out of the detailed national components.

ORCID iDs

Lorena Zardo Trindade https://orcid.org/0000-0001-9646-6996
Tim Goedemé https://orcid.org/0000-0002-2985-4320

Data Availability

The MetaSILC 2015 database is freely available in excel to data users and other interested parties, and can be downloaded from: https://doi.org/10.7910/DVN/TLSZ4S.

Competing interests

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript: Lorena Zardo Trindade and Tim Goedemé

References

Eurostat. 2016. Methodological guidelines and description of EU-SILC target variables. DocSILC065 (2015 operation), European Commission. https://circabc.europa.eu/sd/a/afb4601b-4e5c-4f40-86bb-0c3d0d94aa12/DOCSILC065operation2015VERSION08-08-2016.pdf [Accessed 06.02.2020].

Goedemé T, Zardo Trindade L. 2020a. MetaSILC 2015: A database on the contents and comparability of the EU-SILC income variables [Data file], Antwerp: Herman Deleeck Centre for Social Policy, University of Antwerp & Oxford: Institute for New Economic Thinking, University of Oxford. https://doi.org/10.7910/DVN/TLSZ4S

Goedemé T, Zardo Trindade L. 2020b. MetaSILC 2015: A report on the contents and comparability of the EU-SILC income variables. Oxford: Institute for New Economic Thinking at the Oxford Martin School, University of Oxford and Antwerp: Herman Deleeck Centre for Social Policy, University of Antwerp. DOI: https://doi.org/10.13140/RG.2.2.16292.63368

Zardo Trindade L, Goedemé T. 2020. The comparability of the EU-SILC income variables: review and recommendations. In: *Statistical Working Paper Series, 2020 Edition*. Eurostat, Luxembourg: Publication Office of the European Union. DOI: https://doi.org/10.2785/047001

Appendix

Table A.1 List of EU-SILC income-related variables and those included in MetaSILC 2015

Code	Name
HY010 (*)	Total household gross income
HY020 (*)	Total disposable household income
HY022 (*)	Before social transfers other than old-age and survivor's benefits
HY023(*)	Total disposable household income before social transfers including old-age and survivor's benefits
HY030G/HY030N	Imputed rent
HY040G/HY040N (*)	Income from rental of a property or land
HY090G/HY090N (*)	Interest, dividends, profit from capital investments in unincorporated business
HY050G/HY050N (*)	Family/children related allowances
HY060G/HY060N (*)	Social exclusion not elsewhere classified
HY070G/HY070N (*)	Housing allowances
HY080G/HY080N (*)	Regular inter-household cash transfer received
HY081G/HY081N	Alimonies received (compulsory+ voluntary)
HY100G/HY100N	Interest repayments on mortgage
HY110G/HY110N (*)	Income received by people aged under 16
HY120G/HY120N (*)	Regular taxes on wealth
HY130G/HY130N	Regular inter-household cash transfer paid
HY131G/HY131N	Alimonies paid (compulsory+ voluntary)
HY140G/HY140N (*)	Tax on income and social contributions
HY145N (*)	Repayments/receipts for tax adjustment
HY170G/HY170N	Value of goods produced for own consumption
PY010G/PY010N (*)	Employee cash or near cash income
PY020G/PY020N	Non-cash employee income
PY021G/PY021N (*)	Company car
PY030G	Employer's social insurance contribution
PY031G	Optional employer's social insurance contributions
PY035G/PY035N	Contributions to individual private pension plans
PY050G/PY050N (*)	Cash benefits or losses from self-employment
PY070G/PY070N	Value of goods produced for own consumption
PY080G/PY080N (*)	Pension from individual private plans
PY090G/PY090N (*)	Unemployment benefits
PY100G/PY100N (*)	Old-age benefits
PY110G/PY110N (*)	Survivor's benefits
PY120G/PY120N (*)	Sickness benefits

Code	Name
PY130G/PY130N [*]	Disability benefits
PY140G/PY140N [*]	Education-related allowances
PY200G	Gross monthly earnings for employees

Source = Eurostat (2016a).

Notes: The suffix G refers to Gross; the suffix N refers to values net of taxes and social contributions. (*) Indicates that the variable was analysed in the context of the MetaSILC 2015 report Goedemé and Zardo Trindade, 2020b.