



The  
University  
Of  
Sheffield.

# Introduction to the European Union Statistics on Income and Living Conditions (EU-SILC)

Dr Alvaro Martinez-Perez  
ICOSA Research Associate



# Workshop overview

1. EU-SILC data
2. Data Quality Issues
3. Issues working with EU-SILC
4. Access to the data

# EU-SILC data

- The EU's Statistics on Income and Living Conditions (EU-SILC), launched in 2003 is the first ex-post harmonised micro-data set to provide comprehensive data on incomes and a large number of other social and economic domains across all 28 member states of the enlarged EU (it also covers a number of other countries)

# EU-SILC data

- Eurostat establishes the regulatory framework:
  - Output-harmonised definitions: not same questionnaires but a set of economic and social indicators which should be provided. It is up to each of the statistical offices of the member states how these are to be collected
  - Minimum methodological requirements: probability sampling, fieldwork, etc.
  - Methodological recommendations

# EU-SILC data

- **Two modules**

- Cross-sectional module (2004 onwards): Intergenerational transmission of poverty, Social participation, Housing conditions, Over-indebtedness and financial exclusion, Material deprivation, Intra-household sharing of resources, Intergenerational transmission of disadvantages
- Longitudinal module (2005 onwards): demographics, housing, material deprivation, social exclusion, income, education, health, labour, childcare

- **Measurement units:** Households and individuals

# EU-SILC data

- **Target population**

All private households and their current members residing in the territory of the countries at the time of data collection. All household members are surveyed, but only those aged 16 and more are interviewed

- **Sample size** (minimum size of the overall population surveyed every year):

XS: 130,000 HH and 270,000 individuals aged 16 and more are interviewed

L: ABOUT 100,000 HH and 200,000 individuals aged 16 and more interviewed

# EU-SILC data

- The XS and L files are released separately in the EU-SILC data. Member states are allowed to use different survey instruments to collect XS and L data, and there is no requirement that these data sets be linkable. Indeed, they are not because even when cross-file identifiers to allow linkage are provided by the National Statistical Offices these are changed by Eurostat prior to the public release of the data

# EU-SILC data

- **Data collection:**

Annual XS (households and individuals) (2004 onwards)

Longitudinal: Individual and household-level variables with a minimum of a three year spell permanence (4 waves).

**ROTATING PANEL** (2005 onwards)



The sample is divided into sub-panels. Each sub-panel is retained in the sample for a maximum of four years, and each year one sub-panel is dropped, to be replaced by a new replication



# EU-SILC data

- **Exceptions to the four-years rule:**

- France: nine-year panel

- Norway: eight-year panel

- Luxembourg: 'pure' panel

- **Exceptions to the rule of household surveys:**

- Finland, the Netherlands, Norway, Slovenia, and Sweden have based data collection on administrative registers, using registers to collect several variables, and obtaining other info via interviews with a 'representative person' in the household

# EU-SILC data

- **Participating countries:**

Since 2007. the EU-SILC includes all 27 Member States plus Turkey, Switzerland, Norway and Iceland

Croatia (28 EU MS) will be added in later releases of EU-SILC

2004 includes 13 EU Member States: AT, BE DK, ET, FI, FR, EE, IE, IT, LX, PT, ES, SE (+ Norway and Iceland)

In 2005 and 2006: Add to the former DE, NL, UK and the new EU Member States CY, CZ, HU, LV, LT, PL, SK, SI

# Data quality issues

- **Sampling and data collection:**

DE: Data collected using a mix of quota and random samples. Quota samples should not be used to infer info about the population. No indicator of sample type in the data set

ES & IE use substitutes for non-respondents. Non-respondent substitution undermines the probability nature of the sample. Ideally, substitutes should not be used but the data sets do not contain an indicator to identify them

RECOMMENDATION: Data from these countries should not be used in statistical analyses with the aim of inferring info about national populations

# Data quality issues

- **Sampling and data collection:**

EU-SILC does not report the year in which each sample member was selected. It appears that some sample members of the same rotational group were first interviewed in a different year than others but not possible to identify them

- **Complex sample design indicators:**

In order to obtain correct estimates sample design needs to be taken into account: stratum indicator, cluster indicator and weights

All EU-SILC countries with the exception of DK, SE, IC, implement stratification as part of their sample design , but no stratum indicator is available as part of the EU-SILC data set

# Data quality issues

- **Cluster indicator:**

As clustering affects the precision of estimates if it is not taken into account at the analysis stage, standard errors will be underestimated, and relationships which are not statistically significant may appear to be significant

The EU-SILC does provide info on clustering but in a form which makes it impossible to use in the majority of analyses

The clustering indicator (PSU) should be present for each respondent, for each year in which she appears in the data set, and should represent the cluster of the HH where the person lived at the time of selection

# Data quality issues

- **Weighting:**

Weights are provided with the EU-SILC data. However, the documentation is unclear on whether these are only design weights (correcting for the probability of selection into the sample) or whether they adjust for non-response, and if so, in what countries.

It appears that some countries correct for non-response but they do so in a non-consistent way across countries.

Different treatment of non-response, and especially correcting for it in some countries but not others, may lead to biased cross-country comparisons

# Issues working with EU-SILC

- **Relationships between household members**

EU-SILC does not provide a household grid (a series of variables documenting the relationship between each member of the household and each of the other members)

EU-SILC provides only three variables: the personal identifiers of each individual, his or her spouse or partner, and his or her mother and father, where these are resident in the same household (able to identify which people are living as part of a couple, and/or with their children or parents)

This limitation means that the nature of many relationships cannot be established

# Issues working with EU-SILC

- **The parent/child relationship**

For each individual with one or more parents living in the same household, the mother's and/or father's identifier is supplied. However no distinction is made between biological parents, adoptive parents, foster parents, and step parents

If a full household grid were to be collected making the proper distinctions, this problem would be solved



# Issues working with EU-SILC

- **Issues with the longitudinal module**

The EU-SILC documentation states the rule of following up those sample members aged 16 and over should be re-interviewed if they leave the original household and start living somewhere else. However the implementation of this rule has varied widely from country to country, and has not been particularly comprehensive anywhere.

There are two groups of people who may be particularly affected: young adults and people who divorce or separate

# Issues working with EU-SILC

- **Issues with the longitudinal module**

This means that, although the percentage of individuals and households followed from year to year in the longitudinal sample is not bad, the percentage of individuals followed on leaving their family home is extremely low, making the EU-SILC in its current form unsuitable for analysing transitions for some of the groups of most interest to social scientist: young adults and separating couples

# Issues working with EU-SILC

- **Links between cross-sectional and longitudinal files**

XS and L files cannot be linked which poses a limitation for the variables that are only present in one of the files

A more important limitation is that individuals and households cannot be linked across years in the XS files

When using pooled data from the EU-SILC XS files, some individuals and/or households will be present for only one year, while for others there will be repeated observations. In order to obtain the correct standard errors for estimates, this clustering must be taken into account, but there is no way to determining which are the observations repeated

# Issues working with EU-SILC

- **Income aggregation**

The NSAs gather info about incomes from a large number of sources. However, in the data released by EU-SILC, these incomes are aggregated into a much smaller number of variables

- (+) Increases comparability across countries

- (-) Decreases the level of detail in the data

- (-) As benefit systems differ between countries, the income sources contained in each of the aggregate variables vary between countries, and it is not always clear from the documentation what the components of the aggregate variables are

- (-) Some income components are reported at the household level, not allowing the analyst to distinguish which individuals receive them and how much each receives

# Issues working with EU-SILC

- **Reference period mismatch between income and non-income information**

EU-SILC goal is the collection of definitive data about hh income. It implements this by asking about the total income of hh members during the calendar year prior to the survey interview

Problems:

1. Recall bias: causing measurement error particularly for income components that fluctuate over time

2. Timeliness: key hh income info one year old at the time of collection but up to three years by the time the data is release.  
Implications for the usefulness of EU-SILC for policy purposes

# Issues working with EU-SILC

- **Reference period mismatch between income and non-income information**

3. Temporal mismatch between the income reference period and the current reference period. Almost all other data in EU-SILC refers to the moment in which the interview takes place. This affects the reliability of any analysis focusing on the relationship between income and any other variable

To overcome this problem linking income variables at  $t$  with all other at  $t-1$  would mean to lose one wave which for this short panel is problematic

# Issues working with EU-SILC

- **Net *versus* gross income**

Information on gross incomes has been recorded since 2007:

1. Quality of info is not uniform across countries: data on income components may be collected either gross or net of taxes
2. Different countries apply different approaches to convert net values into gross (v.g. microsimulation, statistical methods, matching with administrative fiscal data)

Each income variable contains a flag indicating the collection method but the quality of this flag info is not uniform across countries

# Access to the data

## Contract

- UoS EU-SILC data received contains the XS (2004-2011) and L (2005-2010)
- Contract signed with Eurostat until 31/12/2015. New releases of data should be made available until then
- After that date the contract needs to be amended to extend its validity



# Access to the data

## Data access

The data is storage in a virtual drive in ICOSS:

*Website* (steps to follow):

<http://www.sheffield.ac.uk/icoss/eu-silc>

1. Sign the confidentiality agreement form
2. Access will be granted (either bring to ICOSS a hard drive with 8GB capacity or you will be given access from your PC to the drive where the data is storage)
3. Please **DO NOT WORK DIRECTLY** with the data on that drive **BUT SAVE IT FIRST** on to your **OWN DIRECTORY** or **EXTERNAL DRIVE**